REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
2/6/08	Final Report	3/15/2004 - 6/30/2007

4. TITLE AND SUBTITLE

(STTR PH II) REAL-TIME DETECTOR OF HUMAN FATIGUE

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-06-C-0008

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Dr. John Stern

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

BIO-BEHAVIOR ANALYSIS
SYSTEMS, LLC

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
USAF, AFRL/NL
AF Office of Scientific
Research
875 Randolph Street, Rm. 3112
Arlington, Virginia 22203

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-SR-AR-TR-08-0103

12. DISTRIBUTION / AVAILABILITY STATEMENT

Distribution A: Approved for Public release

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project seeks to develop superior new methods to monitor human alertness and fatigue in real time, using a suite of unobtrusive sensors and algorithms to assess and predict human behavioral capabilities.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE		29	DM Rocke
					19b. TELEPHONE NUMBER *(include area code)* (530) 752-6999

Contractor: Bio-Behavior Analysis Systems, LLC
Contract No.: FA9550-06-C-0008
Project Title: (STTR Phase II) Real-Time Detector of Human Fatigue
Report Title: Final Technical Report
Date: 15 Feb 2008

REAL-TIME DETECTOR OF HUMAN FATIGUE:

Detecting Lapses in Alertness

Final Technical Report

REAL-TIME DETECTOR OF HUMAN FATIGUE:

Detecting Lapses in Alertness

Final Technical Report

1 General Considerations

1.1 Introduction

Our concern is with the development of software and hardware necessary for the identification of bio-behavioral measures of lapses in alertness. We speak of multiple measures because it is our conviction that no single measure will identify such events for all subjects and because multiple events may more reliably index lapses in alertness. Our focus is on the development of "on line", "real-time" procedures for lapse identification.

Our conceptual model of alertness or the converse, lapses in alertness requires two components. The first component, referred to as "tonic" changes in alertness, involves the slow and probably steady decline in alertness that occurs as subjects become "tired", "fatigued", "bored", etc. by task performance. The second component, referred to as "phasic" changes in alertness, involves the occurrence of momentary performance lapses. The operator returns to his normal level of performance following such a lapse. The two components are not unrelated, we suspect, for example, that an increase in frequency and duration of momentary (phasic) lapses may signal the early phases of tonic alertness decrements. In any case, operators will demonstrate phasic lapses before tonic changes can be detected. Based on our current research effort, we can demonstrate marked individual differences in the speed with which momentary performance lapses evolve. We further note that there is reasonable within subject consistency in such effects.

One further concern is with the development of procedures that can be readily implemented in the "real world". It is our conviction that citizens will not accept procedures requiring attachment of sensors. We thus limit our efforts to measures that can be obtained without the attachment of sensors to the operator. Our current investigations utilize:

a. Video/camera technology to monitor oculomotor information
b. Laser-Doppler Vibrometry (LDV) to capture cardiovascular and muscle activity
c. Pressure sensors installed on equipment operated by the user (In the current context, these pressure sensors were attached to a computer mouse to monitor behavioral variables.)

The following measures have been utilized (and to varying extents validated) with respect to detection of alertness loss.

Behavioral Measures

a. Reaction time; delayed responses
b. Missed signals and false alarms

c. Anticipatory responses – These are response events that occur close in time to the expected stimulus.)

d. Response duration – A little studied measure which we find to be associated with Time-on-Task (ToT) effects.

e. Motor restlessness – This increases with ToT.

Physiological Measures

a. Eye blink and eyelid closure – Including blink closure duration, blink frequency, blink amplitude, and PERCLOS (partial lid closure, Wierwille et al. 1994)

b. Eye movements (principally saccades) – saccade frequency, saccade duration, overshoot and undershoot saccades

c. Timing of blinks with respect to saccades

d. Pupil diameter – Changes in pupil diameter reflect not only light intensity but also reflect aspects of cognition and affect.

e. Pupillary hyppus – slow fluctuations in pupil diameter associated with fatigue

f. Minor head movements – associated with difficulty in information processing

g. Heart period or heart rate – cardiac slowing associated with aspects of information processing

h. Derivative cardiac measures – e.g. left ventricular ejection period, reflection waves, etc.

1.2 Development of the Enhanced Psychomotor Vigilance Task (EPVT)

It was our intent to develop a behavioral task that would, without the introduction of sleep deprivation, demonstrate alterations in performance as a function of ToT. We required a task that would reliably generate performance lapses, where such lapses could be objectively defined, and one that would require gaze shifts. The Psychomotor Vigilance Task (PVT, Dinges & Powel, 1985) partially meets the above requirement with the exception that it does not require gaze shifts and contains little in the way of cognitive activity. It involves presentation of a count-up counter at the center of a computer display. Subjects are instructed to press a key as soon as they see the counter incrementing. The key press stops the counter and the subject's reaction time (RT) is displayed. It is thus a *simple* reaction time task involving *detection* of change. This task is sensitive to sleep deprivation as well as psychoactive medication.

Since in the "real world" people generally are required to search aspects of their environment, and since we are interested in using oculomotor measures to identify lapses in alertness, we altered the PVT by presenting count-up timers at three equidistant locations (in the horizontal plane). This requires the operator to shift gaze to the target location to abstract RT information and makes the new task a *recognition* reaction time paradigm. Donders (1868) has demonstrated that shifting from simple to recognition RT increases response latency. To make the task

cognitively as well as perceptually demanding, we added a second constraint, a *running memory* component. Subjects were required to press the right mouse button following three successive trials where the RT ended in an odd integer. The last integer was under computer control. We refer to our altered version of the PVT as the *Enhanced Psychomotor Vigilance Task (EPVT).*

Similar to previous evaluations of PVT data, we used long latency responses to index attentional lapses. In a 60-minute run of the EPVT, subjects make in excess of 1000 RT responses. (The absolute number of events presented and responses given depends on the speed with which subjects respond. Subjects who respond more rapidly experience a larger number of stimuli and provide a larger number of responses.) We provisionally accepted the 50 longest RT events to index attentional lapses.

We evaluated ToT effects by averaging responses over successive five or ten minute periods. With respect to the identification of bio-behavioral measures reflecting attentional lapses, we divided these into "tonic" and "phasic" measures. Tonic changes were reflected in changes in average performance and changes in bio-behavioral measures over successive five or ten minute periods. Phasic measures were those associated with the occurrence of long latency responses. Our major concern was with the identification of bio-behavioral events associated with momentary (phasic) attentional loss.

Because performance on the running memory component of the EPVT in our first study was far from perfect, we developed software to allow for feedback to the subject about performance on that aspect of the task. The feedback utilized the words "correct", "wrong", and "miss" to inform the subject about response accuracy on the running memory component of the task. This information was provided aurally 0.5 seconds following occurrence of the first two types of events ("correct" and "wrong") and 3.5 seconds following the presentation of the third odd integer when the operator missed making a correct running memory task response.

1.3 Development of Hardware for Bio-Behavioral Measurements

In addition to developing the EPVT, we believed it desirable to develop more sensitive measures of responding on the part of the operator. It has long been known that common computer operating systems (e.g. Microsoft Windows™) and computer mice are not designed to provide highly accurate timing information about mouse related events (Segalowitz & Graves, 1990, Beringer, 1992). We thus set about the task of installing sensors on the mouse that would provide more information about how the operator went about the task of pressing the mouse buttons. We evaluated a number of possible techniques and singled out the use of pressure sensors as the most reliable procedure for obtaining the desired precision for monitoring button presses. The initial study for this project was conducted with the sensors mounted directly on the face of the left and right mouse buttons. However, we found that some subjects, especially those with narrow fingers, were able to response (as reflected by mouse switch closure) without activating the pressure sensors. To reduce this problem in later studies, we mounted pressure sensors under the left and right front portion of the mouse as well as at the back of the mouse. With this arrangement there are still occasions in which we do not get an output from the pressure sensors though a mouse switch is closed and reopened. However, these events occur relatively infrequently.

1.4 Development of Software

Since our ultimate concern is with the developing procedures to reduce periods of attentional loss by detecting lapses in alertness and providing feedback to operators and supervisors, we developed computer software for the on-line analysis of the bio-behavioral measures of interest. Our initial effort in this direction was the development of software to provide feedback to the operator about performance of the running memory task. This feedback mechanism is described above.

More significantly, we have developed an *Application Framework for Alertness Monitoring Applications*. The framework is designed to allow for accepting data from a variety of data sources and to allow for the creation of alertness monitoring applications via "plugging" in event detection and pattern recognition algorithms. This software is described more fully in Section 4 below.

2 Studies Conducted

The following studies where conducted as part of this project.

2.1 Evaluation of Measures Reflecting Tonic and Phasic Alertness Lapses

Twenty college age subjects (10 male and 10 female) provided the basic data demonstrating the utility of the EPVT with respect to producing alertness lapses in non-sleep-deprived subjects. Subjects performed the task for 60 minutes.

2.2 Evaluation of Cognitive Complexity's Effect on Reaction Time (RT)

Twelve subjects performed the EPVT as well as separately performing on the RT component of the EPVT (without the running memory portion of task) to determine whether the memory load imposed by the running memory task affected RT.

2.3 Evaluation of EPVT Performance of Sleep Disturbed Individuals

Fifteen HIV positive patients suffering from sleep disturbance performed the EPVT for 30 minutes each.

2.4 Evaluation of Repeat Reliability of EPVT Performance

Four subjects performed the task on 3-4 occasions to allow us to evaluate within subject consistency of performance. These subjects received feedback on their performance of the running memory component of the task.

2.5 Evaluation of Head Movements in Acquisition of Information

Six subjects participated in a study evaluating the impact of perceptual changes on the use of head movements in the acquisition of information. This study used an on-screen reading task instead of the EPVT.

2.6 Demonstration of the Effects of Caffeine on EPVT Performance

One subject was used in a demonstration of the effects of caffeine deprivation and caffeine intake on EPVT performance and on bio-behavioral measures.

3 Results

3.1 Evaluation of Measures Reflecting Tonic and Phasic Alertness Lapses

The following analyses were performed on data gather as part of the study described in Section 2.1 above.

3.1.1 Simple RT ToT Effects

Mean reaction time was calculated for all subjects for successive five-minute periods. Table 1 depicts these averages for every subject.

Mean Reaction Time per five minute period										
min	**f10**	**f12**	**f13**	**f14**	**f17**	**f18**	**f20**	**f21**	**f22**	**f23**
00-05	0.396	0.356	0.357	0.490	0.373	0.300	0.382	0.418	0.439	0.319
05-10	0.366	0.359	0.362	0.833	0.394	0.304	0.369	0.445	0.562	0.324
10-15	0.428	0.358	0.352	1.582	0.408	0.305	0.360	0.485	0.556	0.328
15-20	0.496	0.331	0.363	1.442	0.451	0.314	0.424	0.450	0.638	0.338
20-25	0.525	0.344	0.381	1.127	0.486	0.317	0.415	0.462	0.587	0.346
25-30	0.602	0.351	0.379	0.610	0.457	0.317	0.407	0.468	0.475	0.344
30-35	0.648	0.371	0.411	1.047	0.478	0.314	0.420	0.436	0.409	0.352
35-40	0.612	0.379	0.502	0.928	0.752	0.325	0.421	0.421	0.448	0.349
40-45	0.710	0.371	0.504	1.283	0.576	0.325	0.434	0.428	0.475	0.343
45-50	0.930	0.357	0.461	1.016	0.702	0.320	0.441	0.453	0.466	0.358
50-55	0.749	0.361	0.523	0.832	0.957	0.323	0.499	0.433	0.502	0.362
55-60	0.589	0.372	0.545	0.445	0.787	0.336	0.486	0.435	0.425	0.362
min	**m10**	**m11**	**m12**	**m13**	**m14**	**m15**	**m16**	**m17**	**m18**	**m22**
00-05	0.344	0.369	0.321	0.527	0.377	0.291	0.359	0.320	0.342	0.412
05-10	0.360	0.403	0.351	0.822	0.352	0.325	0.361	0.320	0.363	0.446
10-15	0.358	0.438	0.353	0.941	0.387	0.312	0.343	0.361	0.378	0.416
15-20	0.350	0.488	0.356	0.799	0.392	0.308	0.375	0.354	0.401	0.410
20-25	0.351	0.512	0.413	0.735	0.368	0.319	0.355	0.475	0.412	0.408
25-30	0.380	0.609	0.390	0.925	0.367	0.353	0.389	0.517	0.464	0.410
30-35	0.387	0.579	0.402	0.504	0.388	0.325	0.435	0.450	0.472	0.425
35-40	0.392	0.500	0.440	0.409	0.403	0.338	0.570	0.574	0.518	0.437
40-45	0.402	0.577	0.481	0.404	0.394	0.359	0.629	0.534	0.663	0.439
45-50	0.417	0.644	0.452	0.422	0.434	0.421	0.644	0.615	0.786	0.480
50-55	0.450	0.616	0.537	0.462	0.419	0.400	0.439	0.576	0.803	0.484
55-60	0.405	0.614	0.513	0.417	0.405	0.354	0.724	0.548	0.899	0.537

Table 1: Mean Reaction Times for Successive 5-minute Periods

Figure 1 presents the RT data averaged across all subjects.

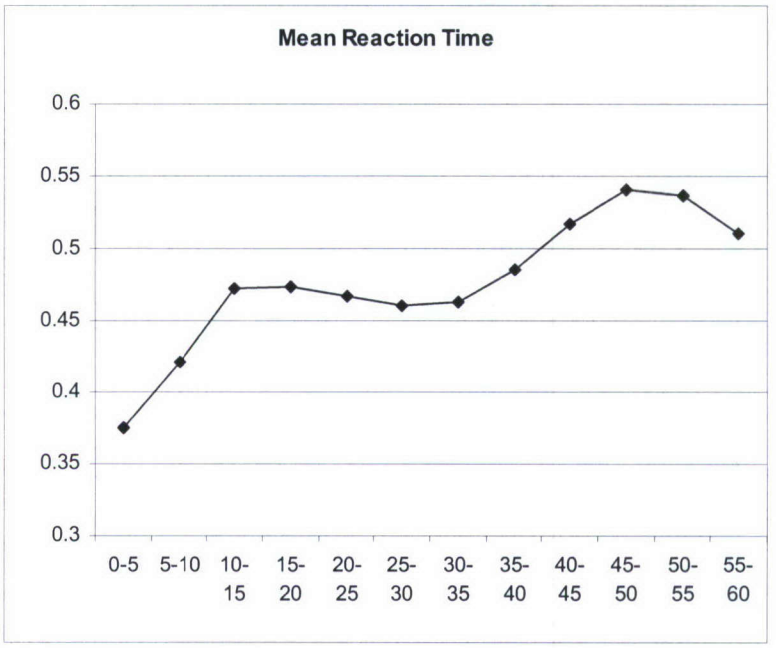


Figure 1: Mean Reaction Time Averaged Across Subjects

As is apparent from Table 1 above, there was considerable variability in RT across subjects. This variability may lead one to wonder how representative the average RT across 20 subjects is of the individual subject. That is, are there subjects who demonstrate the pattern seen in Figure 1? Figure 2 and Figure 3 present the RT data (averaged over 10-minute periods) for the female and male subjects respectively. It is readily apparent that the average presented in Figure 1 is not representative of the individual. Most subjects demonstrate an increase in RT over time, but the degree of the increase is markedly different across individuals. Notably, there are also subjects who do not fit this pattern. Subjects F14 and M13, for example, develop increases in RT early in task performance and improve over the last half of the task.

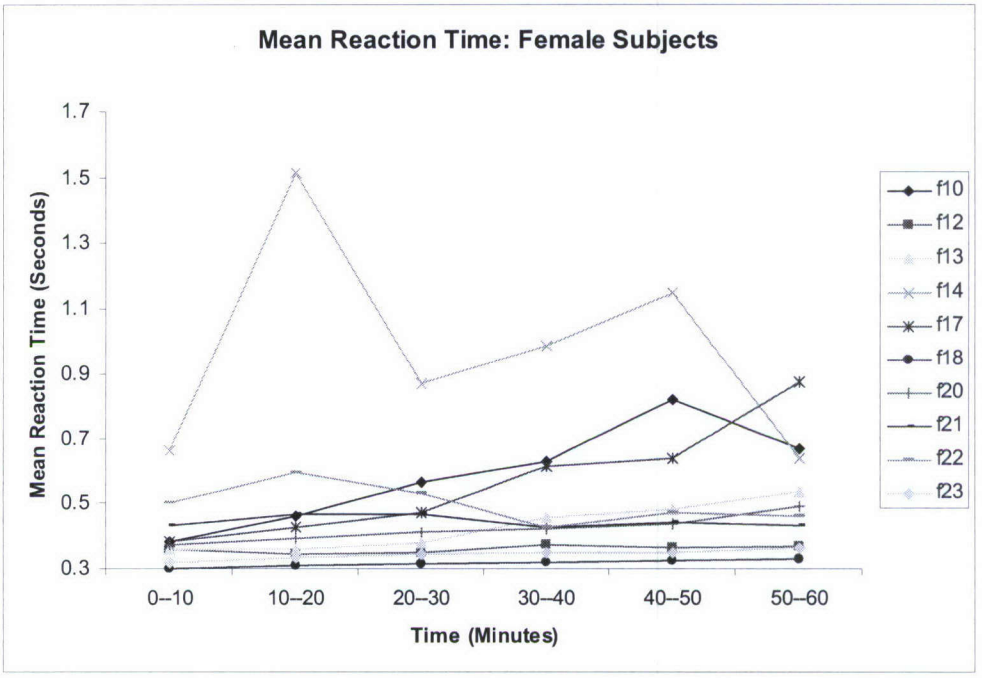


Figure 2: Mean Reaction Time for Individual Female Subjects

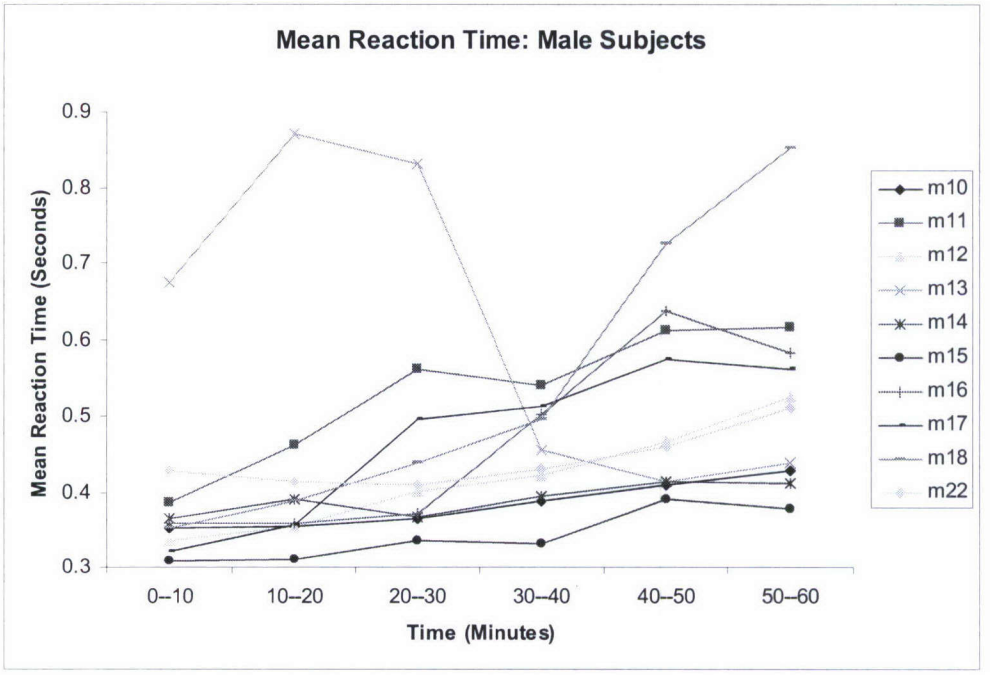


Figure 3: Mean Reaction Time for Individual Male Subjects

3.1.2 Performance on the Running Memory Component of the Task

The running memory component of the EPVT required subjects to make a right mouse click response following a sequence of three trials on which RT ended in an odd integer. The occurrence of such events was under experimenter control by assigning responsibility for the value of the last integer to the computer.[1] The system was programmed so that approximately 40 to 50 events required a right click response. We say approximately since the total number of trials generated in the 60-minute period depended on the subject's RT.

After subjects had read the instructions for the task, they were given a 5-minute training session. During this training period, the experimenter observed the subject's performance paying particular attention to the performance of the running memory portion of the task. This assured the experimenter that the subject understood the task instructions. Performance on the running memory portion of the task (a.k.a. the Memory Task or MT) was evaluated with respect to Hits, False Alarms, and Misses. To determine whether there was a ToT effect, the data was divided into successive 10-minute periods. This is depicted in Figure 4. We see little change in the frequency of correct responses (Hits) or in the frequency of combined False Alarms and Misses (labeled Mistakes in the figure) as a function of ToT.

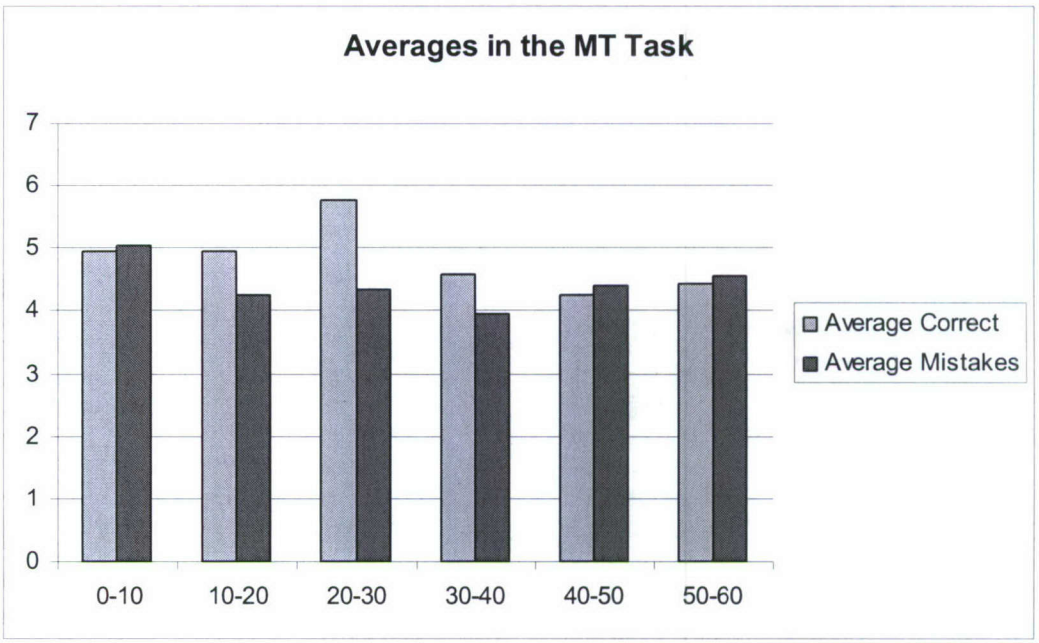


Figure 4: Average Number of Correct and Mistaken Memory Task (MT) Responses

[1] To clarify: The subject's reaction time was presented as a 4-digit value representing the reaction time in milliseconds. The last digit of this reaction time value was not necessarily an accurate representation of the subject's reaction time. If the subject had taken 1598 milliseconds to respond as measured by the computer presenting the task, and this was a trial that was predetermined by the experimenter to end in an odd digit, then the actual reaction time presented to the subject might be 1593. Similarly, if the computer determined reaction time was 1649 and the trial was one that was to end with an even digit, 1648 might be presented to the subject as her reaction time.

These results indicate that performance was far from perfect for most subjects. The two best performing subjects (one male, one female) had 48 hits during the session. The male subject with 48 hits had 3 false alarms and 1 miss during the session. The female subject with 48 hits had 1 false alarm and 2 misses during the session. At the other extreme the poorest performing male had 3 hits, 20 false alarms, and 46 misses. The poorest performing female had 8 hits, 14 false alarms, and 40 misses. For the male subject with only 3 hits, those 3 hits occurred in a relative narrow time window between minute 5 and minute 15 of the task. In the case of both the male and female subjects with the poorest performance, the distribution of combined false alarms and misses did not discriminate between the initial and final 30 minutes of task performance. This was also true of the group of subjects in general. On the other hand, missed signals were considerably more frequent in the last 30-minute period.

Though subjects had been trained on the EPVT, and we assured ourselves that they understood the instructions, their performance on this component of the task was far from perfect. Table 2 below identifies the hit rate for each of the 20 subjects as well as the "t" test comparison between male and female subjects.

hit rate - % correct R.M. task			
Female	64.6	Male	59.2
	87.8		64.6
	83.3		37.5
	39.0		47.8
	89.4		68.0
	96.0		06.1
	69.4		49.0
	81.3		39.6
	16.7		53.1
	75.5		81.3
t-Test: Two-Sample Assuming Equal Variances			
	Female	*Male*	
Mean	70.300	50.620	
Variance	614.949	422.746	
Observations	10.000	10.000	
Pooled Variance	518.848		
Hypothesized Mean Difference	0.000		
df	18.000		
t Stat	1.932		
P(T<=t) one-tail	0.035		
t Critical one-tail	1.734		
P(T<=t) two-tail	0.069		
t Critical two-tail	2.101		

Table 2: Hit Rate Statistics

Accuracy ranged from 6 to 96%. Though the "t" test analysis discriminating between male and female subjects was not statistically significant ($p<0.069$), we conducted a second analysis utilizing the Mann-Whitney "U" test. Performance was ranked from highest to poorest. The sum of ranks for female subjects was 76, that for males 114. These differences are significant at the 0.01 level. Female subjects are somewhat slower in responding to the reaction time component of the task. Their better performance on the running memory component may thus be attributable to a speed/accuracy trade-off.

We can thus infer that adding the running memory component to the simple RT component enhances task difficulty. This was born out when we ran a group of older subjects who were both HIV positive and sleep disturbed. These subjects had much more difficulty understanding the memory task instructions than was true of our college students and made even more errors on the running memory task. As is described below, there are significant differences in the RT between performing the EPVT and performing only the recognition RT component of the task. This is another indicant that enhancing information processing demands increases RT, even on the timing of the decision/recognition that an incrementing count-up timer has started.

For the RT component of the task, subjects received immediate feedback because the count-up timer stopped when they responded. Unlike the RT component of the task, the running memory component had no such feedback. Because of this lack of feedback on the running memory component of the task and the general difficulty of that part of the task, we decided to develop a second version of the task, one in which subjects also received feedback about their performance on the running memory task. We programmed the computer to provide voice feedback using the words: "Correct" for a hit, "Wrong" for a false alarm, and "Miss" for a missed running memory task response. This change markedly improves accuracy on this component of the task.

3.1.3 Long Latency Responses – There Relationship to Lapses in Alertness

A number of procedures can be used to identify performance "outliers" or long latency responses. The approach taken by Dinges et al. was to use the mean and standard deviation measures for their identification. We, in this initial investigation, sampled the 50 longest latency responses for each subject. Since each subject made in excess of 1000 responses during the 60 minutes of task performance, this essentially sampled 5% of the events. For comparison purposes we also sampled the 50 fastest responses. It was our assumption that if there was a ToT effect, we would find a distribution skewed to the right for long latency responses and one skewed to the left for short latency responses. Figure 5 and Figure 6 depict these results. We do not need statistical evaluation to demonstrate that the above suggestion about shape of the distributions is verified. For the distribution of fastest responses, we see a steady decline with an asymptotic level for the last 30 minutes of task performance. For the 50 slowest responses we see an increase from the initial 5-minute period to 5-10 minute period, a continued increase to the 10-15 minute period, an essentially flat level from the 10-15 minute period through minutes 30-35, and a steady increase for the next 20 minutes of task performance. We are thus reasonably assured that 60 minutes of EPVT performance leads to an increase in lapses in alertness as reflected in long latency simple reaction time responses. It is interesting to note that the distribution of long latency responses mirrors the mean reaction time data. There we see an increase in R.T. for the

initial 15 minutes of task performance, an asymptotic level for the next 20 minutes and an increase for the following 20 minutes.

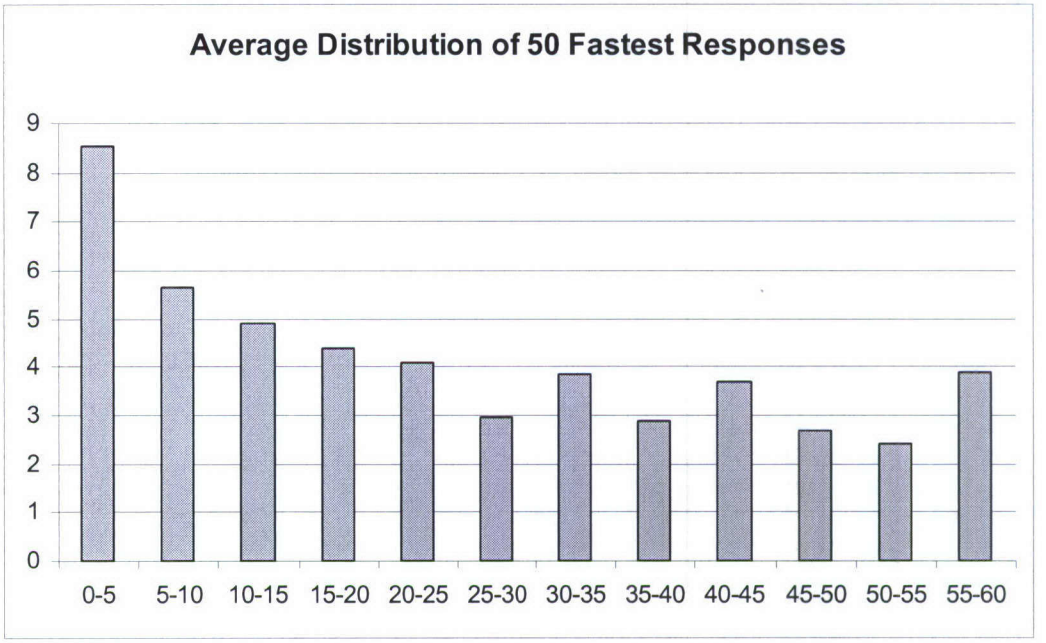


Figure 5: Distribution of 50 Fastest Responses

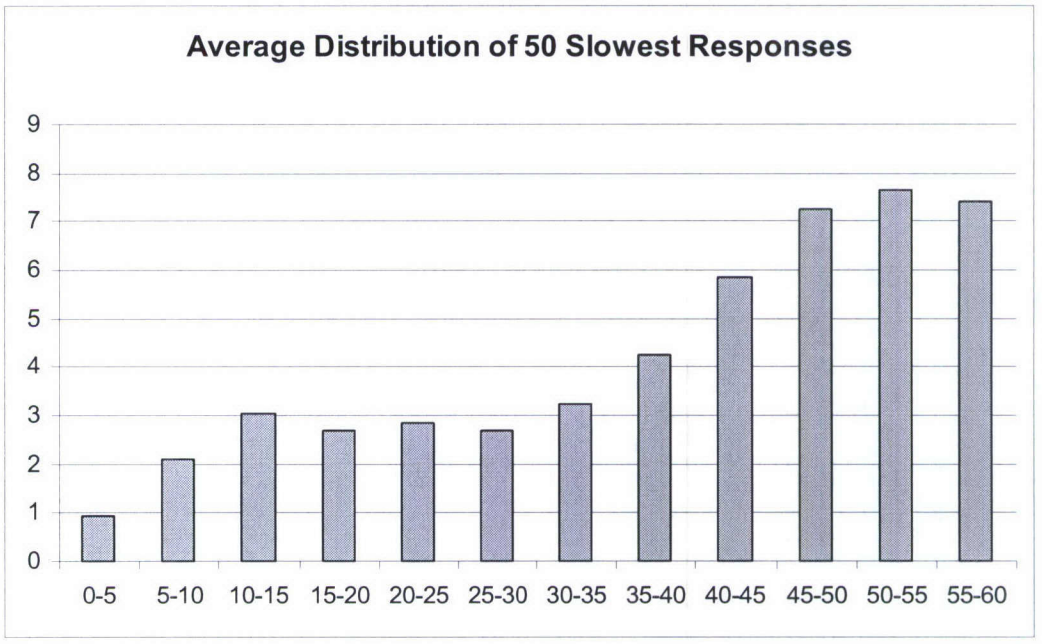


Figure 6: Distribution of 50 Slowest Responses

Another procedure for characterizing long latency responses is to evaluate the difference between the mean and median RT. The arithmetic mean is more affected by outlying values than is true of the median. Thus, if there is an increase in long latency responses as a function of ToT, the curves representing RT as measured by these two procedures should pull apart. Figure 7 depicts the average mean and median RT for this data set. Figure 8 shows the same data "normalized" by using the initial five minutes as the covariate.

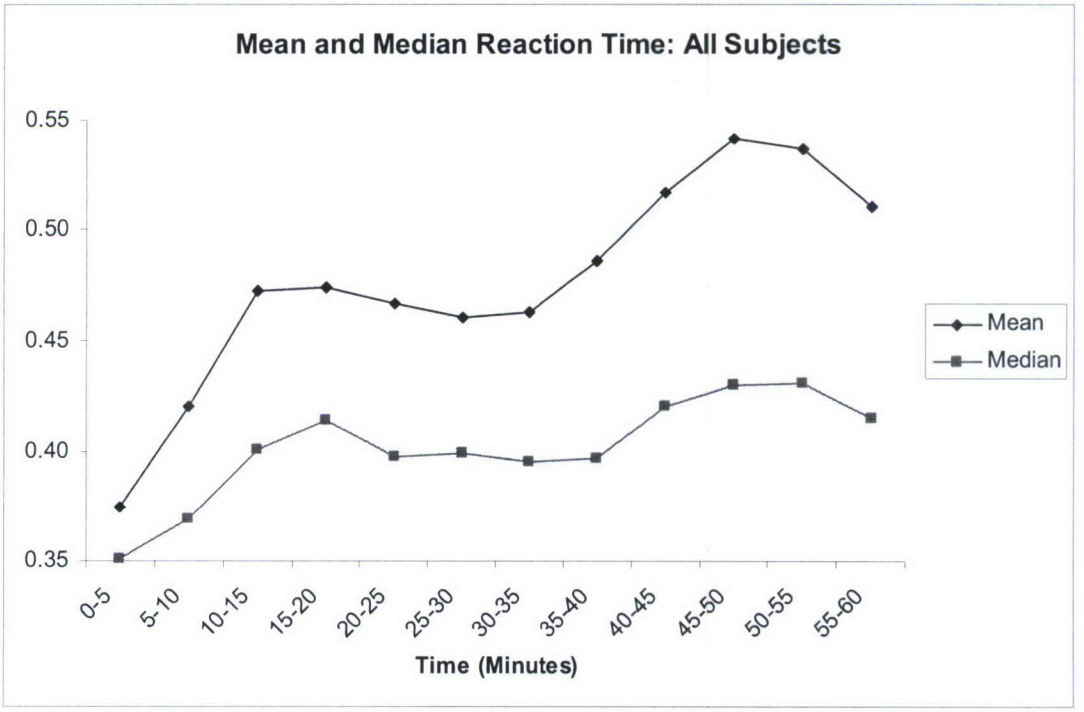


Figure 7: Mean and Median Reaction Time

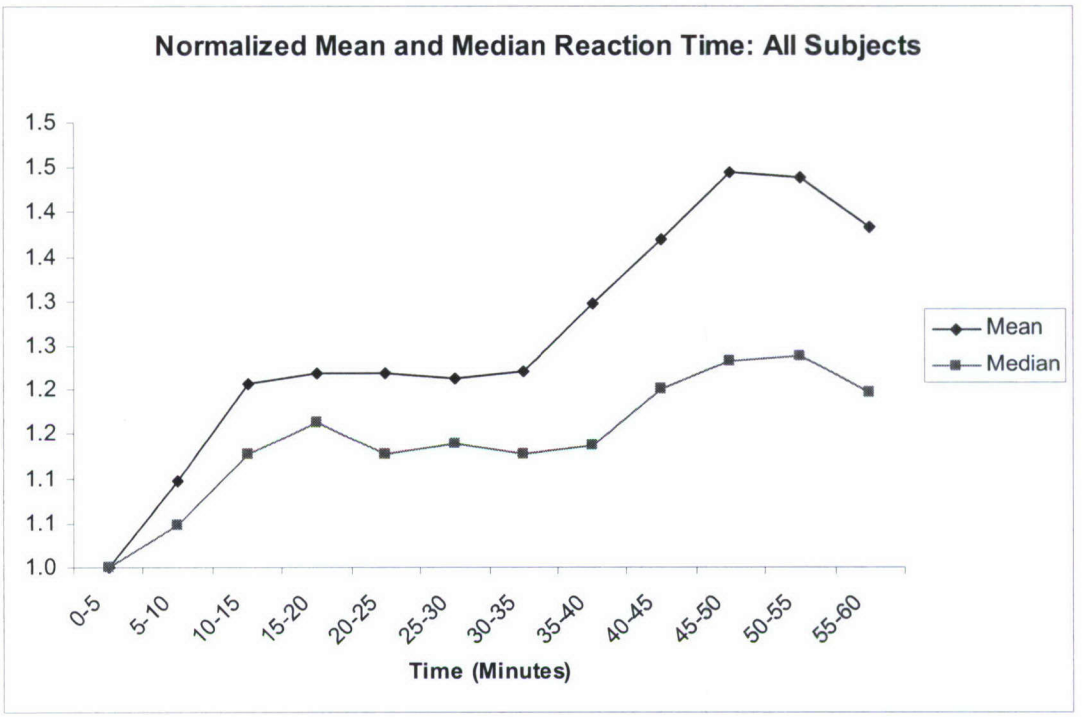


Figure 8: Normalized Mean and Median Reaction Times

3.1.4 Evaluation of Pressure Sensors as Determiners of RT

As described above we developed two versions of pressure sensor enhanced mice. The first, used in study 1, placed the pressure sensing elements directly on the left and right mouse switches. The second version, used in later studies placed the pressure sensors underneath the mouse with one for the left, one for the right and one at the back of the mouse.

Since data abstraction of the pressure response in study 1 was conducted manually, we only evaluated the first five subjects for whom pressure sensor responses were available. The difference between RT as measured with the output from the mouse (operating in the Windows™ environment) and from the pressure sensor placed on the left mouse button was sampled for a minimum of 50 RT events selected randomly over the 60-minute experiment. In all cases the pressure sensor identified RT occurred earlier than that obtained from the mouse.

Table 3 presents the results of this analysis. Each column in the table represents a different subject for which this analysis was performed. The values in the table are the differences between the mouse RT and the pressure-based RT (Mouse RT minus Pressure RT) and are given in seconds. There are consistent differences between the two data sets with the pressure sensor output preceding the mouse output. The difference in mean values for the 5 subjects is narrow, ranging from 0.055 seconds to 0.064 seconds. Variability in this measure is also in a reasonably narrow range.

Mean	0.064	0.055	0.058	0.059	0.064
S.D.	0.012	0.021	0.018	0.021	0.013
Median	0.064	0.059	0.061	0.058	0.062
Minimum	0.044	0.007	0.033	0.020	0.047
Maximum	0.089	0.093	0.110	0.121	0.101

Table 3: Reaction Time Differences between Mouse and Pressure Sensor Output

We have developed software allowing us to monitor pressure sensor outputs and have applied this in the analysis reported below. This is still work in progress.

3.1.4.1 Are the pressure sensors a reliable (and valid?) procedure for monitoring "reaction time"?

"Reaction time" is generally defined as the time taken between detecting a stimulus requiring a response and the enactment of that response. In the past, enactment has generally been measured by switch closure. The time required between initiating switch depression and full closure is presumed to be a constant. If we are interested in the time it takes to react, it is better to monitor initiation of a movement rather than its completion. Some experimenters have done this by monitoring muscle activity, the precursor to the actual movement. Since our concern is with measuring bio-behavioral signals without attachment of sensors to the operator, we decided to investigate the utility of monitoring pressure exerted on the mouse switches to identify response initiation.

Why might a pressure sensor be a more reliable and sensitive indicator of reaction time than mouse output or switch output for monitoring reaction?

a. A pressure sensor captures the initiation, rather than the completion of a response. Thus the output from the pressure sensor is a better reflector of "detection" of stimuli.

b. A pressure sensor captures pressure exerted not only when a mouse switch closure identifies a response, but at other times as well. Visser et al. (2004) demonstrated that "mental pressure" significantly increased output from the pressure sensor. Ulrich, Mattes & Miller (1999) identify variables such as stimulus and response probability, stimulus duration and intensity, and time pressure as affecting response force. We have found that some mouse-identified responses are unusually late because the mouse did not capture the initial response. That is, there was an earlier response identified by the pressure sensors that for some reason was not registered by the mouse.[2]

c. Multiple pressure sensors can capture the dynamics of the movement. For example, having a pressure sensor at the back and the front of the mouse allows one to determine whether, associated with a switch closure, there is a release of pressure on the back of the mouse preceding the initiation of increased pressure on the front of the mouse. We are also interested in what are referred to as "overflow" movements. For

[2] Perhaps the pressure applied was not enough to cause the mouse switch to close. Perhaps some other factor involving the mechanics of the mouse or the workings of the operating system caused the mouse button press to not be registered.

example extra pressure applied to the right mouse button concurrent with a depressing of the left mouse button or an increase in overall pressure on the mouse. Additional examples of the "dynamics" of the movement that can be captured are: the amount of pressure exerted to activate a response, the time between response initiation and peak pressure, the time from peak pressure to response completion.

d. As has been mentioned earlier in this document, sampling of output from the mouse switch using the Windows™ operating system is unreliable (Plant, Hammond, & Whitehouse, 2003).

In addition to using the pressure sensors as a more reliable and sensitive indicator of reaction time, there are other, non-reaction-time, measures that can be obtained using the pressure sensors. In particular, we were interested in capturing a measure of "restlessness" on the part of the operator. This restlessness is expected to increase with ToT and be reflected in changes in pressure (increases, decreases, spikes, and drops in pressure) on the mouse. Baker (1960) reports on monitoring gross body activity by placing microswitches under the subject's pivotally mounted chair. He found significant increase in such activity as a function of ToT but did not find it useful in identifying the level of vigilance during task performance.

Our initial procedure placed pressure sensor elements on the left and right mouse switch as well as on the left side where the subject's thumb was most likely to rest. Output from the system was unreliable. Unreliability was produced principally by the subject's ability to place the responding finger outside of the location of the pressure sensor.

Based on this experience we re-instrumented a mouse with pressure sensors located on the underside of the mouse, one underneath the left mouse switch, a second underneath the right mouse switch, and a third sensor at the back of the mouse. The sensors have small buttons glued to them directly under the sensing element. Thus, pressure exerted on the mouse can be monitored. We have been experimenting with this device, including the development of algorithms to capture the movements of interest. We are in the process of developing another version of this mouse, one that will allow us to capitalize on the operational amplifier that is part of the sensor package and increase voltage output from the mouse.

As described below we have used the mouse while subjects were performing the EPVT and, though there are still some problems to be solved, we think that the output from the combined sensors will allow us to reliably monitor activity of interest. We plan to make some changes in the physical set-up of our experiment to maximize reliable outputs from the mouse. For example, at the current time the mouse is placed on an inclined plane, this increases the likelihood of increased pressure on the back portion of the mouse. The solution is to place the mouse on a horizontal and flat surface. We find that there are still occasions where the left pressure sensor is not or minimally activated though we obtain a button press response. When there is no response from that sensor, we still obtain outputs from one or the other of the remaining sensors. Since we are primarily interested in the timing of response initiation, we can use that output to monitor response latency.

Results

For the present, we have limited analysis to two components of the response as obtained from the pressure sensors:

1. An increase or decrease in pressure that marks the first response activity after a stimulus and
2. The recovery (return to baseline pressure levels) from that response.

Left mouse button response initiation frequently follows the pattern:

a. Release of pressure on the back of the mouse (sometimes preceded by a very brief and small amplitude increase in pressure.)

b. Initiation of pressure on the front left of the mouse. Initiation of pressure change on the left front pressure sensor precedes activation of the mouse switch. The average time between the two events is approximately 50 milliseconds.

c. Output from the rear pressure sensor is more complex than that from the left front pressure sensor. A common pattern is: Release, Press, Release, Press. When this pattern is seen, the second Release, Press pair usually follows the termination of the left front sensor response (see Figure 11 below).

The figures below illustrate several pressure response scenarios that are representative of what is seen when the pressure signals are clean. In each of these figures, seven channels of data are shown. From top to bottom in the figures, the channels are:

Channel 1. Eye/Gaze Position, Horizontal Plane, Left Eye

Channel 2. Eye/Gaze Position, Horizontal Plane, Right Eye

Channel 3. Back Pressure Sensor Output

Channel 4. Left Front Pressure Sensor Output

Channel 5. Stimulus Presentation to the Left of the Display

Channel 6. Stimulus Presentation to the Right of the Display[3]

Channel 7. Response initiation as identified by output from the mouse itself, not the pressure sensors

If both channels 5 and 6 have displacements, this indicates that a stimulus (count-up timer) was presented at the center of the display. Channel 7 only provides an indication of the onset of a mouse response. That is, the initial displacement on the channel indicates when the mouse/computer combination reported that the mouse switch had been activated. The duration of the "pulse" on channel 7 is not reflective of how long the subject pressed the mouse button. The duration of the displacement (or "pulse") on the two stimulus channels indicates whether the last digit of the reaction time presented was even or odd. Long duration pulses (~200 milliseconds) on a stimulus channel indicate a stimulus ending with an odd digit. Short duration pulses (~100

[3] If both channel 5 and channel 6 have displacements, this indicates that a stimulus was presented at the center of the display.

milliseconds) on a stimulus channel indicate a stimulus ending with an even digit. Data was sampled at 1000 Hz.

A number of other channels of data are available but not shown in the following figures. This include output from the right front pressure sensor, gaze position information for each eye in the vertical plane, minor head movements as inferred from the change in position of the center of the eyeball, pupil diameter, and response termination as identified by output from the mouse.

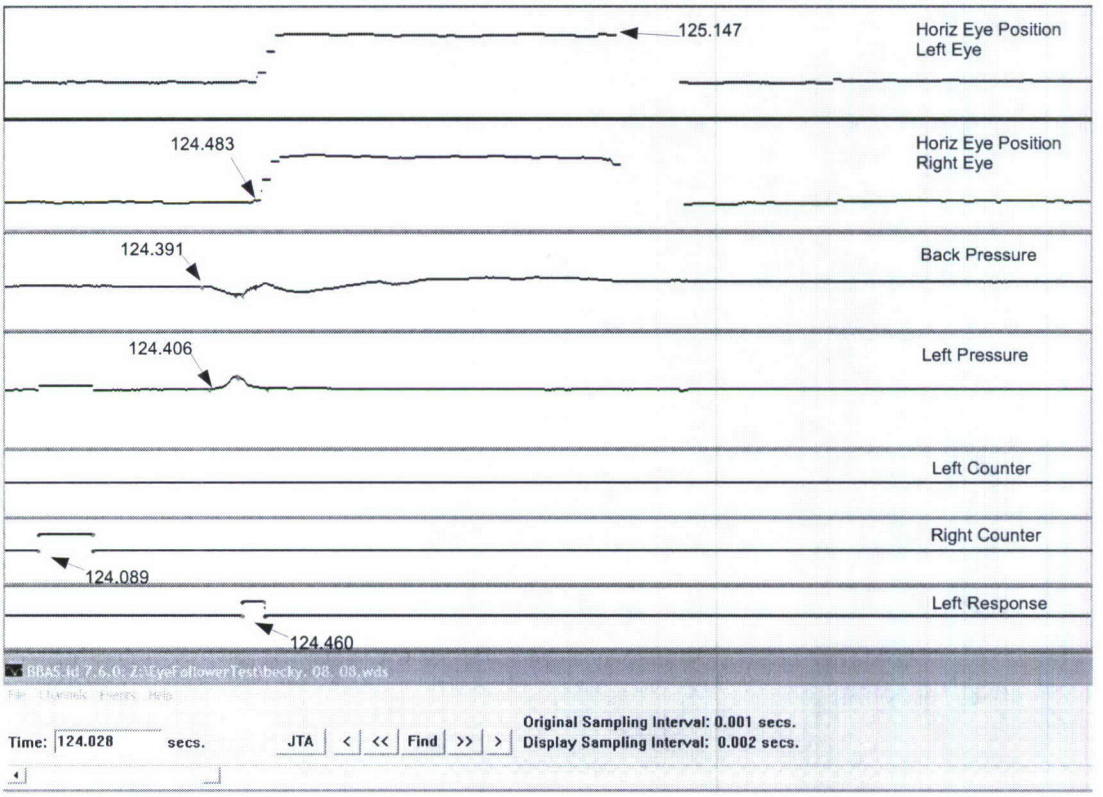


Figure 9

The following can be noted about Figure 9 above:

1. A stimulus (a count-up timer starting on the right side of the display) is presented at 124.089 seconds from the beginning of the recorded data. (The actual experimental run started approximately 8 seconds prior to this stimulus presentation.)
2. The earliest response activity is seen in the release of pressure on the back pressure sensor at 124.391 seconds. The RT as calculated using this initial release of pressure is 0.302 seconds.
3. A pressure change on the left front pressure sensor is observed at 124.406 seconds. This is 15 milliseconds after the release of pressure on the back pressure sensor.

4. The mouse, operating system, and computer recognize that the mouse button switch has been closed at 124.460 seconds. Thus the reaction time, as calculated from stimulus presentation to identification of mouse button switch closure is 124.460 - 124.089 = 0.371 seconds.

5. Gaze shift to the stimulus location occurs at second 124.483. This is 77 milliseconds after the initiation of pressure on the left pressure sensor.

6. The gaze begins its return to the central location at second 125.147. There is data loss in the gaze channels associated with this gaze return. The data loss occurs because the pupil is occluded by an eye blink. The duration of the data loss is 0.118 seconds, a duration that is appropriate for loss attributable to a blink.

7. In the pressure channels we see the following patterns of activity. (All time values are given in seconds.)

	Initiation	**Termination**	**Duration**
Left Front			
Increase	124.406	124.458	0.052
Decrease	124.460	124.510	0.050
Rear			
Decrease	124.391	124.450	0.059
Increase	124.464	124.506	0.042
Decrease	124.506	124.574	0.068
Increase	124.574	124.718	0.144

Table 4: Pressure Response Patterns for Figure 9

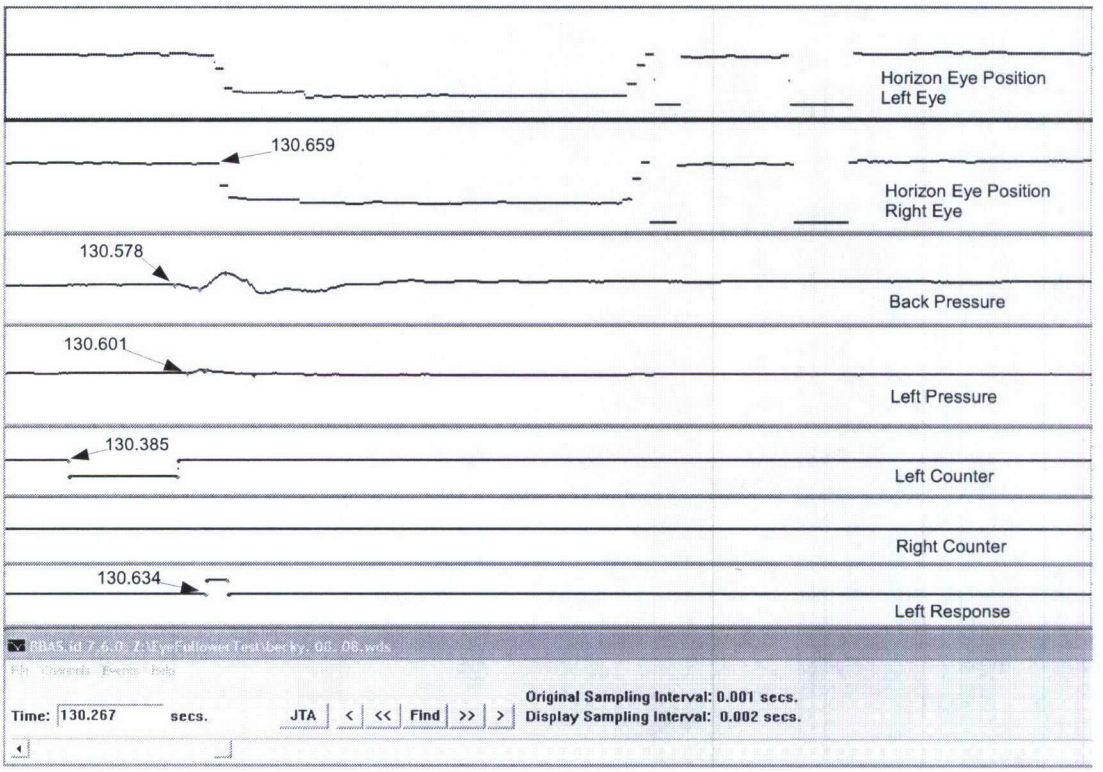


Figure 10

The following can be noted about Figure 10 above:

1. A stimulus (a count-up timer starting on the left side of the display) is presented at 130.385 seconds.
2. The earliest response activity is seen in the release of pressure on the back pressure sensor at 130.578 seconds. The RT as calculated using this initial release of pressure is 0.193 seconds.
3. A pressure change on the left front pressure sensor starts at 130.601 seconds. This is 0.023 seconds after initiation of the release of pressure on the back pressure sensor.
4. The mouse button switch is recognized as having been closed at 130.634 seconds. The RT using this indicator is 130.634 – 130.385 = 0.249 seconds.
5. A gaze shift to the location of the stimulus initiates at 130.659 seconds. The gaze shift returning the eyes to the center location initiates at 131.413 seconds. This is following by two blinks. Unlike the previous example, the first of these two blinks does not occur during the eye movement, but follows the termination of the gaze shift by 0.014 seconds.

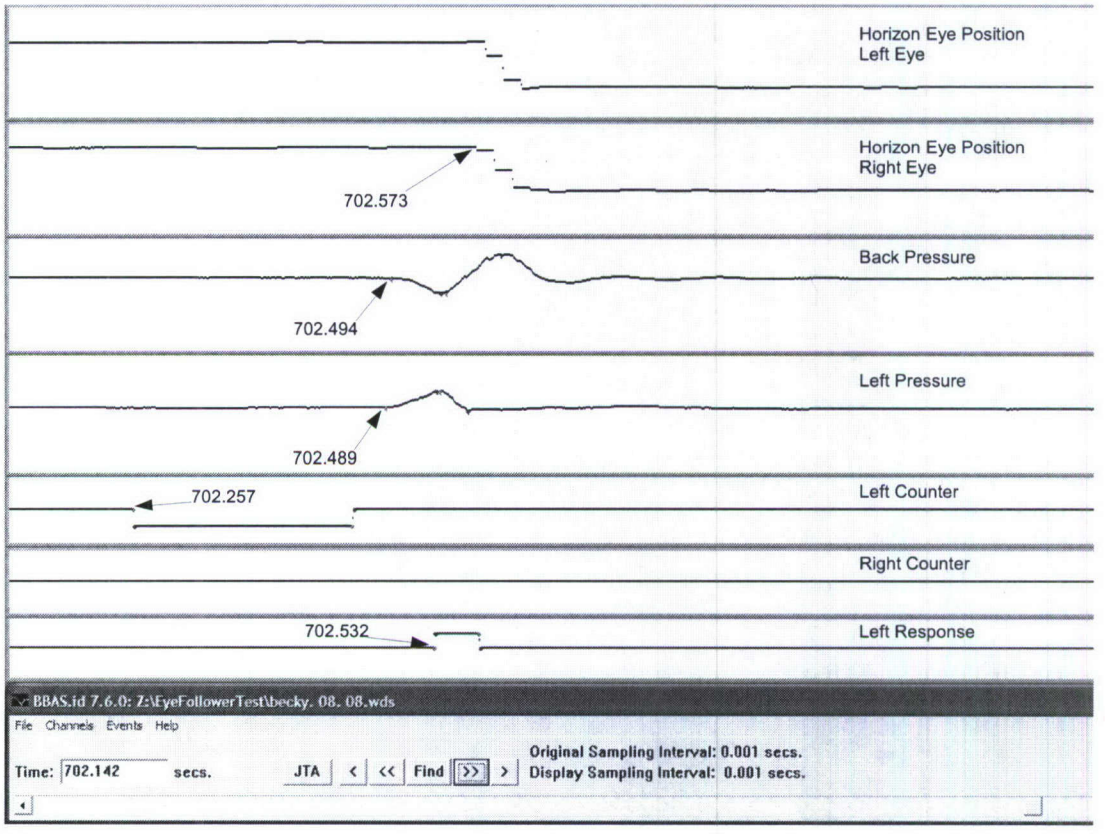


Figure 11

The following can be noted about Figure 11 above:

1. First it should be noted that unlike Figure 9 and Figure 10, this figure shows only one second of data. This second is stretched out horizontally in the figure so that the channels in this figure have greater temporal resolution than the previous two figures.
2. A stimulus is presented at 702.257 seconds.
3. The earliest response activity is seen in an increase in pressure detected by the left front sensor at 702.489 seconds. The RT as calculated using this initial increase in pressure is 0.232 seconds.
4. The initiation of pressure release on the back sensor is 0.005 seconds after pressure initiation on the front left sensor.
5. The RT calculated using the mouse switch closure indication is 702.532 - 702.257 = 0.275 seconds.
6. Gaze shift to the location f the stimulus is initiated at 702.573 seconds. This is 0.0.84 seconds after the initiation of pressure on the left front pressure sensor.
7. Notice the Release, Press, Release, Press pattern in the back mouse pressure channel.

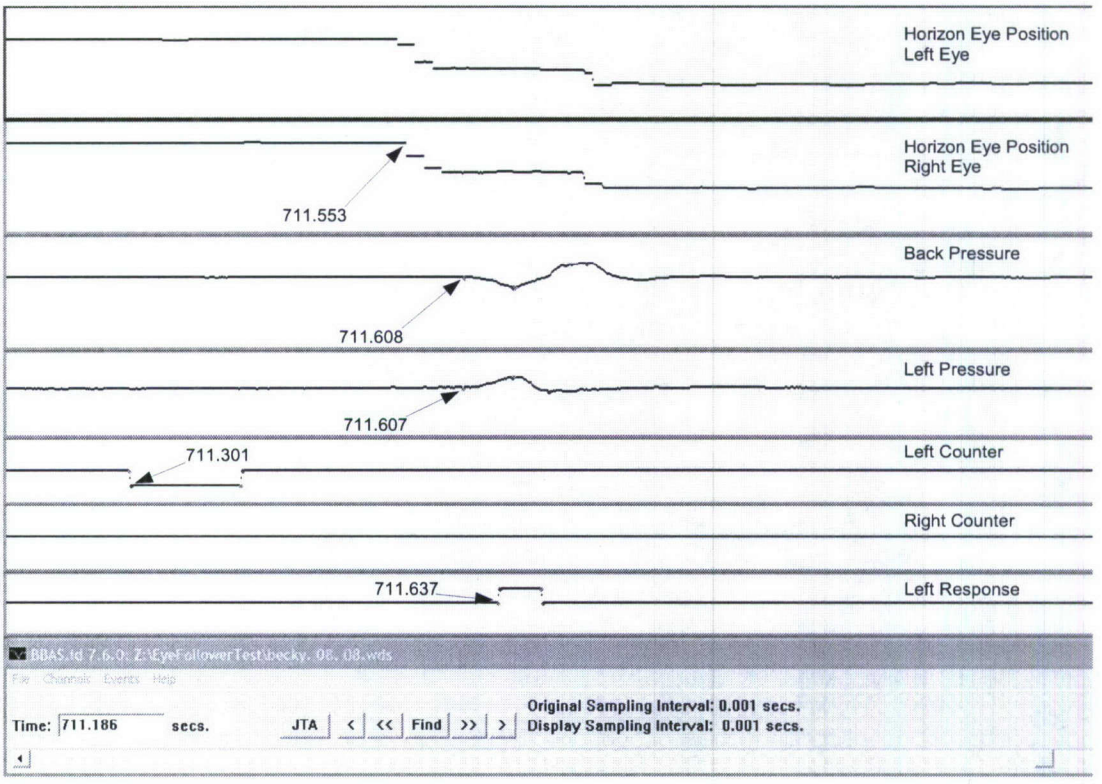


Figure 12

The following can be noted about Figure 12 above:

1. A stimulus is presented at 711.301 seconds.
2. The earliest response activity is from the left pressure sensor at 711.607. RT = 0.306 seconds.
3. Pressure on the back sensor is release at the same time a pressure is exerted on the left front sensor.
4. RT calculated using the mouse output (initiation at 711.637) is 0.336 seconds.
5. Gaze shift to stimulus location initiated at 711.553 seconds. Note that in this case the initiation of the eye gaze shift precedes the change in pressure from the pressure sensors. It is difficult to see in the figure, but it appears that the right eye initiates the movement before the left eye.

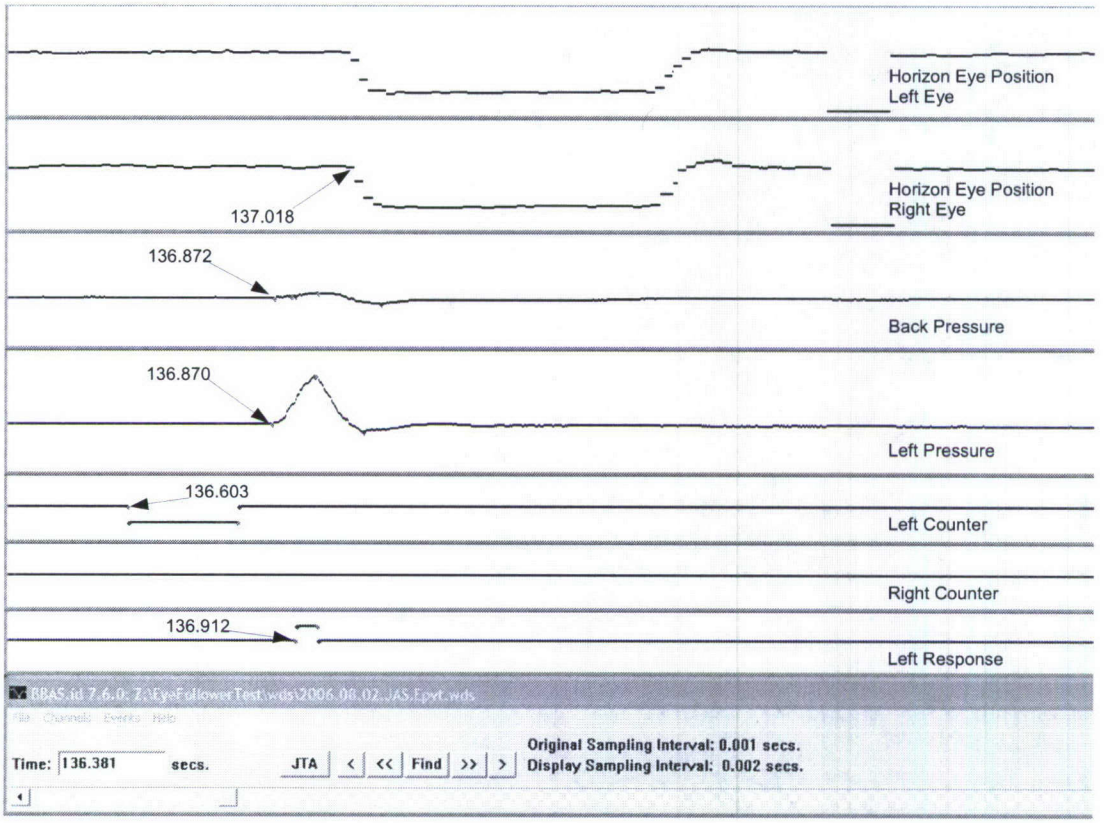


Figure 13

The following can be noted about Figure 13 above:

1. A stimulus is presented at 136.603 seconds. (The actual experimental run started approximately 13 seconds prior to this stimulus presentation.)
2. Earliest response activity is from the left pressure sensor at 136.870 seconds. RT = 0.267 seconds.
3. Pressure is exerted on the back sensor 0.002 seconds later. Notice that for this subject, unlike the previous examples, there does not appear to be any initial pressure release on the back of the mouse prior to the pressure increase on the left front pressure sensor.
4. Using the output from the mouse (Channel 7) we get a RT of 0.309 seconds.
5. Gaze shift to stimulus location initiates at 137.018 seconds, shortly before the termination of the pressure response. Notice that the gaze return is slower than the gaze shift to the target location and that the blink follows the response termination by approximately 0.200 seconds.

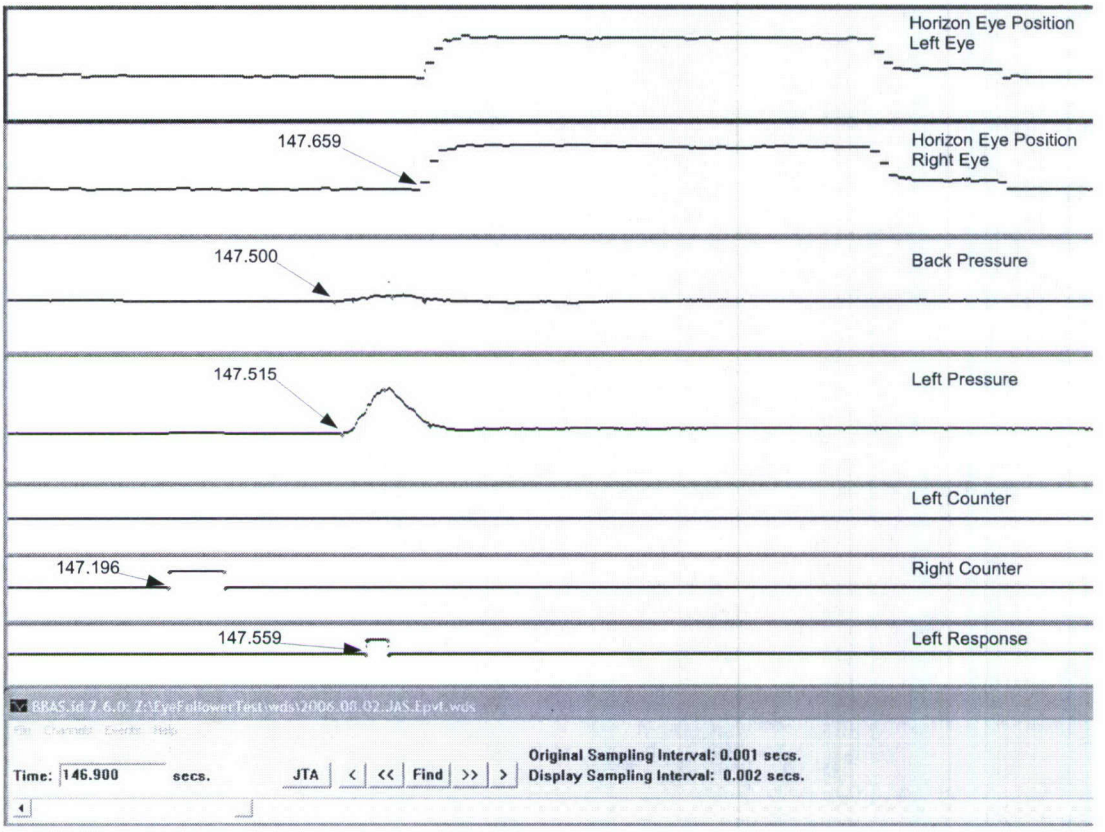


Figure 14

The following can be noted about Figure 14 above:

1. A stimulus is presented at 147.196 seconds.
2. The earliest response activity is an increase in pressure on the back pressure sensor at 147.500 seconds. RT = 0.304 seconds.
3. Left front pressure sensor increase starts 0.015 seconds after increase in pressure on the back pressure sensor.
4. RT calculated using mouse output is 147.559 – 147.196 = 0.363 seconds.
5. Gaze shift is initiated at 147.659 seconds. This is shortly before termination of the pressure response.

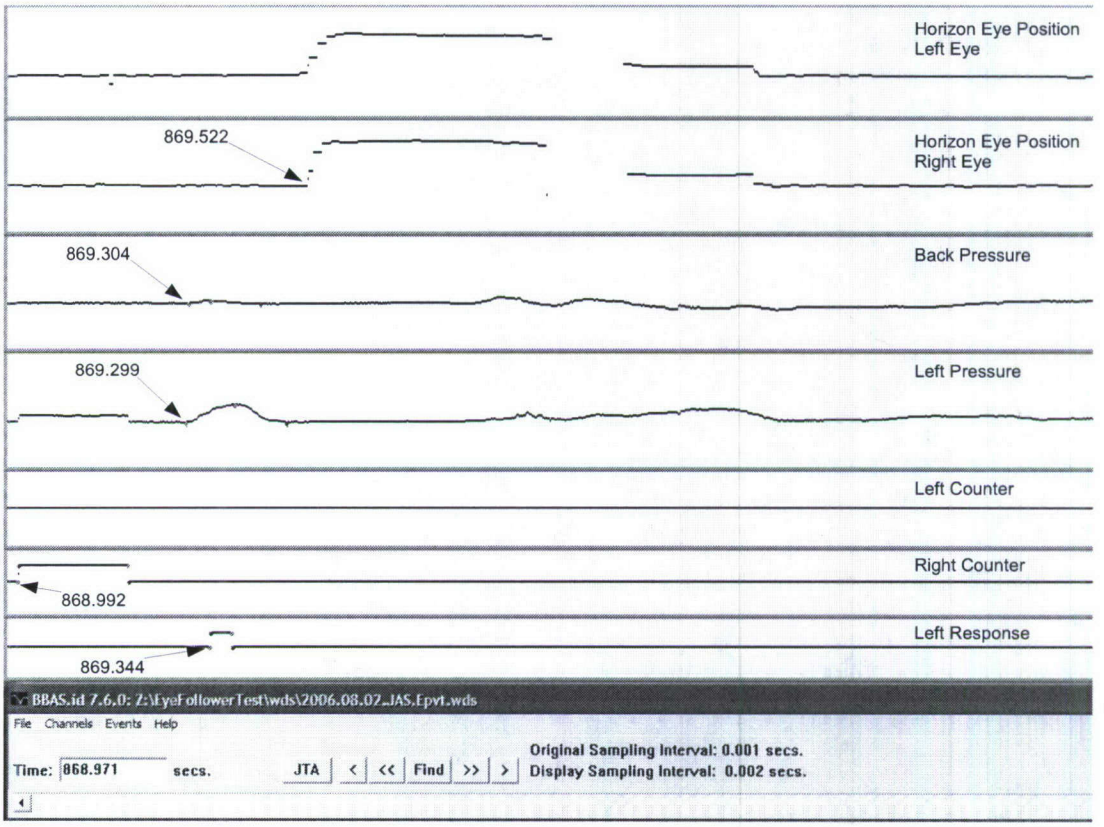


Figure 15

The following can be noted about Figure 15 above:

1. Stimulus onset is at 868.992 seconds.
2. Earliest response activity is from the left pressure sensor. There is an increase in pressure at 869.299 seconds. This yields an RT of 0.307 seconds.
3. There is an increase in pressure on the back pressure sensor that starts 0.005 seconds after the increase in pressure on the left front pressure sensor.
4. RT calculated using the mouse output is 869.344 – 868.992 = 0.352 seconds.
5. The gaze shift to the stimulus location (target) starts at 869.522 seconds. This follows the termination of manual responses. A blink occurs concurrently with the gaze shift back to the center of the display.

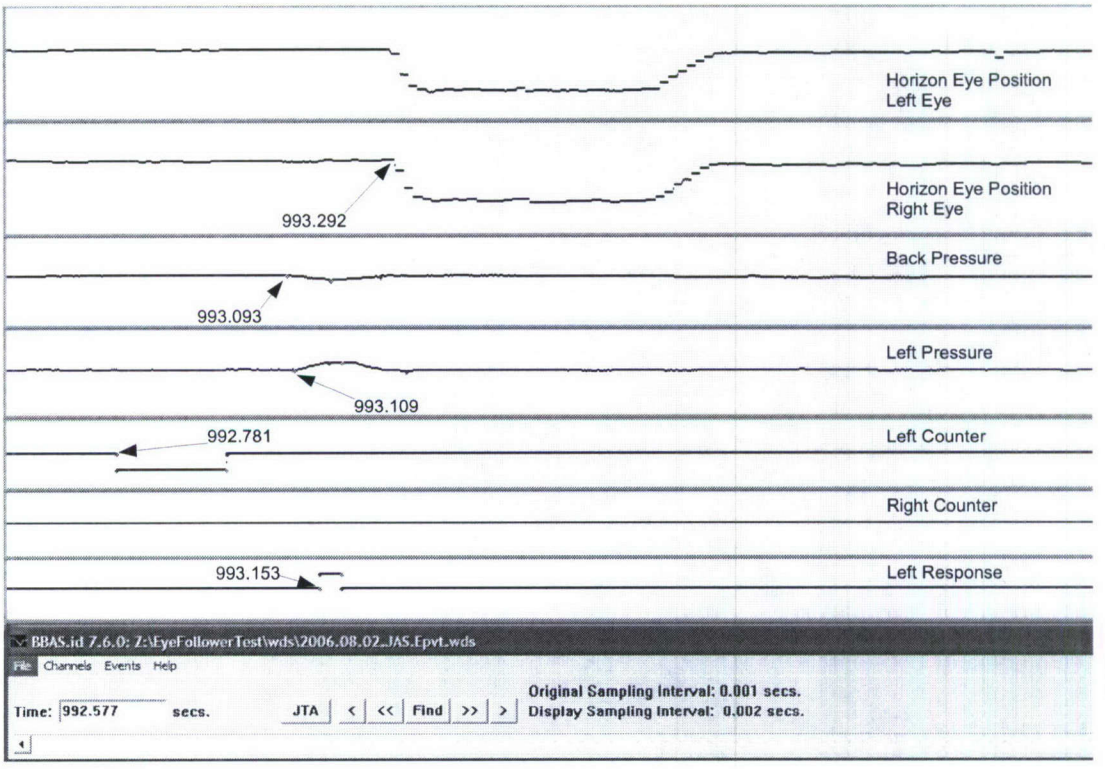


Figure 16

The following can be noted about Figure 16 above:

1. Stimulus onset is at 992.781 seconds.
2. Earliest response activity is from the back pressure sensor. It is a release of pressure starting at 993.093 seconds. RT = 0.312 seconds.
3. The left front pressure sensor registers an increase in pressure starting 0.016 seconds after the release of pressure on the back sensor is initiated.
4. RT calculated using the mouse output is 993.153 - 992.781 = 0.372 seconds
5. The gaze shift to the target location is initiated after the termination of all manual responses.
6. Note that the return saccade is considerably slower than the to-target saccade. The to-target saccade has a duration of 0.056 seconds; the return saccade has a duration of 0.130 seconds.

General Comments about the previous 8 examples

For these 8 example events, the average difference between the mouse-based RT and the pressure sensor based RT was 53 milliseconds with the pressure sensor responses initiated preceding the response as recorded by the mouse.

There appears to be a fairly consistent way of responding on the part of individual subjects. The first four of these eight examples are from the same female subject. She demonstrates a consistent pattern of pressure changes on the back sensor with a release that either precedes or coincides with an increase in pressure on the left front pressure sensor. She also demonstrates a relatively consistent pattern of pressure changes on just the back pressure sensor. There is a release of pressure, followed by an increase in pressure, followed by another release and increase pair.

The second four of these examples are from the same male subject. The pressure pattern for this subject appears to consist of an increase in pressure on the left front sensor with a concurrent increase or no change at all in pressure on the back pressure sensor.

We believe that differences in hand size and placement on the mouse are responsible for some of these differences. The mouse was positioned on an inclined surface during these runs. This may also provide some explanation for these differences.

3.1.5 Response Duration as an Indicator of Alertness Lapses

We have found little in the literature dealing with response duration as applied to simple behavioral measures such as RT. We hypothesized that not only RT would be affected by ToT manipulations but that response duration might also be affected by such a manipulation. We thus, using the mouse based response duration measure, evaluated whether there were any systematic changes associated with ToT. Since there were reasonable differences in response duration (RD) across subjects, we present the data both as mean RD per 5-minute period as well as "normalized" data, where the initial 5 minutes were taken as the covariate and successive 5 minute periods expressed as a ratio of the initial period. Thus if RD increases over time the ratio should be greater than 1.0 and increase over successive periods. Table 5 presents the mean values and Table 6 the "normalized results

	00-05	05-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
F10	0.106	0.123	0.119	0.131	0.128	0.129	0.129	0.132	0.137	0.128	0.133	0.131
F12	0.124	0.119	0.111	0.119	0.124	0.120	0.118	0.114	0.112	0.116	0.114	0.116
F13	0.096	0.119	0.126	0.122	0.127	0.123	0.123	0.120	0.128	0.137	0.141	0.142
F14	0.129	0.187	0.345	0.214	0.184	0.175	0.219	0.348	0.300	0.243	0.305	0.435
F17	0.130	0.140	0.144	0.147	0.150	0.150	0.142	0.147	0.154	0.161	0.163	0.153
F18	0.062	0.060	0.063	0.076	0.089	0.077	0.076	0.082	0.096	0.106	0.087	0.072
F20	0.131	0.127	0.121	0.130	0.132	0.134	0.141	0.139	0.143	0.135	0.131	0.131
F21	0.083	0.089	0.092	0.079	0.089	0.085	0.086	0.081	0.087	0.081	0.089	0.091
F22	0.130	0.148	0.145	0.147	0.150	0.144	0.130	0.148	0.148	0.161	0.133	0.162
M10	0.085	0.096	0.094	0.094	0.094	0.090	0.097	0.098	0.089	0.105	0.109	0.107
M11	0.129	0.127	0.116	0.112	0.120	0.124	0.120	0.123	0.127	0.128	0.129	0.123
M12	0.087	0.096	0.097	0.099	0.100	0.105	0.113	0.124	0.117	0.134	0.137	0.111
M13	0.182	0.235	0.234	0.193	0.212	0.358	0.164	0.143	0.145	0.149	0.148	0.142
M14	0.089	0.105	0.100	0.095	0.096	0.108	0.103	0.113	0.102	0.123	0.130	0.124
M15	0.075	0.070	0.066	0.072	0.066	0.073	0.066	0.062	0.064	0.073	0.075	0.061
M16	0.099	0.107	0.099	0.104	0.125	0.125	0.124	0.125	0.118	0.116	0.118	0.119
M17	0.134	0.154	0.161	0.131	0.131	0.097	0.097	0.103	0.112	0.122	0.120	0.123
M18	0.072	0.077	0.074	0.086	0.073	0.076	0.084	0.095	0.121	0.125	0.147	0.139
M20	0.125	0.125	0.121	0.140	0.137	0.138	0.133	0.131	0.133	0.129	0.135	0.136
Mean	1.068	1.191	1.254	1.182	1.194	1.253	1.166	1.247	1.245	1.264	1.311	1.350

Table 5: Mean Response Duration

	00-05	05-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
F10	1.000	1.153	1.114	1.231	1.231	1.208	1.211	1.240	1.288	1.201	1.250	1.231
F12	1.000	0.958	0.896	0.960	0.960	0.970	0.948	0.916	0.904	0.935	0.920	0.938
F13	1.000	1.239	1.314	1.271	1.271	1.284	1.286	1.253	1.335	1.428	1.466	1.484
F14	1.000	1.450	2.673	1.657	1.657	1.356	1.699	2.693	2.323	1.883	2.360	3.373
F17	1.000	1.075	1.109	1.125	1.149	1.149	1.091	1.131	1.179	1.237	1.252	1.176
F18	1.000	0.965	1.016	1.227	1.430	1.233	1.215	1.317	1.545	1.701	1.406	1.153
F20	1.000	0.969	0.928	0.995	1.008	1.026	1.081	1.067	1.094	1.031	1.005	1.006
F21	1.000	1.065	1.098	0.952	0.952	1.024	1.036	0.971	1.042	0.972	1.072	1.086
F22	1.000	1.138	1.118	1.132	1.132	1.109	1.002	1.139	1.137	1.239	1.027	1.245
M10	1.000	1.130	1.105	1.115	1.115	1.066	1.140	1.156	1.052	1.245	1.288	1.258
M11	1.000	0.981	0.899	0.869	0.869	0.964	0.928	0.957	0.988	0.992	1.002	0.955
M12	1.000	1.109	1.114	1.143	1.143	1.210	1.302	1.424	1.346	1.545	1.576	1.275
M13	1.000	1.296	1.288	1.063	1.063	1.970	0.901	0.788	0.797	0.818	0.813	0.782
M14	1.000	1.184	1.127	1.071	1.071	1.214	1.158	1.277	1.152	1.387	1.470	1.393
M15	1.000	0.931	0.889	0.965	0.965	0.976	0.889	0.827	0.852	0.978	1.004	0.821
M16	1.000	1.081	1.000	1.054	1.054	1.274	1.255	1.273	1.194	1.182	1.201	1.210
M17	1.000	1.151	1.200	0.978	0.978	0.723	0.726	0.769	0.835	0.908	0.897	0.919
M18	1.000	1.061	1.027	1.194	1.194	1.051	1.157	1.312	1.682	1.731	2.042	1.932
M20	1.000	0.999	0.964	1.117	1.117	1.102	1.060	1.043	1.061	1.032	1.079	1.083
Mean	1.000	1.115	1.174	1.107	1.107	1.173	1.092	1.168	1.166	1.184	1.228	1.264

Table 6: Normalized Mean Response Duration

Looking at the above results, one gets the impression that the increase in RD as a function of ToT is, at best, modest, a 26% increase over the 60 minute period. What is equally apparent is that there are subjects who do not show the increase (F12, F18, F20, F21, M13, M15, and M17). Excluding these subjects, the pattern of increasing RD becomes more robust, namely an increase of approximately 50%. These results are depicted in Table 7 and Table 8.

Thus, 30% of subjects do not fit the pattern of increasing RD over time. However, the majority of subjects fit the pattern. RD may be a useful measure reflecting tonic changes in alertness.

	00-05	05-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
F10	0.106	0.123	0.119	0.131	0.128	0.129	0.129	0.132	0.137	0.128	0.133	0.131
F13	0.096	0.119	0.126	0.122	0.127	0.123	0.123	0.120	0.128	0.137	0.141	0.142
F14	0.129	0.187	0.345	0.214	0.184	0.175	0.219	0.348	0.300	0.243	0.305	0.435
F17	0.130	0.140	0.144	0.147	0.150	0.150	0.142	0.147	0.154	0.161	0.163	0.153
F22	0.130	0.148	0.145	0.147	0.150	0.144	0.130	0.148	0.148	0.161	0.133	0.162
M10	0.085	0.096	0.094	0.094	0.094	0.090	0.097	0.098	0.089	0.105	0.109	0.107
M11	0.129	0.127	0.116	0.112	0.120	0.124	0.120	0.123	0.127	0.128	0.129	0.123
M12	0.087	0.096	0.097	0.099	0.100	0.105	0.113	0.124	0.117	0.134	0.137	0.111
M14	0.089	0.105	0.100	0.095	0.096	0.108	0.103	0.113	0.102	0.123	0.130	0.124
M16	0.099	0.107	0.099	0.104	0.125	0.125	0.124	0.125	0.118	0.116	0.118	0.119
M18	0.072	0.077	0.074	0.086	0.073	0.076	0.084	0.095	0.121	0.125	0.147	0.139
M20	0.125	0.125	0.121	0.140	0.137	0.138	0.133	0.131	0.133	0.129	0.135	0.136
sum	1.277	1.449	1.579	1.491	1.483	1.487	1.516	1.704	1.674	1.691	1.781	1.882
mean	0.106	0.121	0.132	0.124	0.124	0.124	0.126	0.142	0.139	0.141	0.148	0.157

Table 7: Mean Response Duration (Selected Subjects)

	00-05	05-10	10-15	15-20	20-25	25-30	30-35	35-40	40-45	45-50	50-55	55-60
F10	1.000	1.153	1.114	1.231	1.231	1.208	1.211	1.240	1.288	1.201	1.250	1.231
F13	1.000	1.239	1.314	1.271	1.271	1.284	1.286	1.253	1.335	1.428	1.466	1.484
F14	1.000	1.450	2.673	1.657	1.657	1.356	1.699	2.693	2.323	1.883	2.360	3.373
F17	1.000	1.075	1.109	1.125	1.125	1.149	1.091	1.131	1.179	1.237	1.252	1.176
F22	1.000	1.138	1.118	1.132	1.132	1.109	1.002	1.139	1.137	1.239	1.027	1.245
M10	1.000	1.130	1.105	1.115	1.115	1.066	1.140	1.156	1.052	1.245	1.288	1.258
M11	1.000	0.981	0.899	0.869	0.869	0.964	0.928	0.957	0.988	0.992	1.002	0.955
M12	1.000	1.109	1.114	1.143	1.143	1.210	1.302	1.424	1.346	1.545	1.576	1.275
M14	1.000	1.184	1.127	1.071	1.071	1.214	1.158	1.277	1.152	1.387	1.470	1.393
M16	1.000	1.081	1.000	1.054	1.054	1.274	1.255	1.273	1.194	1.182	1.201	1.210
M18	1.000	1.061	1.027	1.194	1.194	1.051	1.157	1.312	1.682	1.731	2.042	1.932
M20	1.000	0.999	0.964	1.117	1.117	1.102	1.060	1.043	1.061	1.032	1.079	1.083
sum		13.600	14.565	13.979	13.979	13.987	14.289	15.898	15.736	16.102	17.013	17.614
mean		1.133	1.214	1.165	1.165	1.166	1.191	1.325	1.311	1.342	1.418	1.468

Table 8: Normalized Mean Response Duration (Selected Subjects)

We next asked whether there was a relationship between RT and RD. We evaluated response duration for the 50 longest and shortest RT. With the exception of 4 subjects (F20, M15, M17 and M22), the average RD was longer for long latency as compared to short latency RT. Mean duration for short latency responses was 0.121 and for long latency responses 0.167 (N=20, df

19, t=2.190, p<0.04). The correlation between RD for short and long latency responses was 0.741.

When one correlates RT and RD for the individual for successive 10-minute data periods, the correlations are generally low. See Table 9 below. Thus the pattern of RD development is different from that associated with changes in RT as a function of ToT.

	00-10	10-20	20-30	30-40	40-50	50-60
F10	-0.097	0.131	0.103	0.207	0.185	0.229
F12	0.013	0.251	0.115	0.196	0.213	0.056
F13	0.120	0.004	-0.016	-0.012	-0.072	0.119
F17	0.039	0.184	0.065	0.249	0.195	0.121
F18	0.026	0.022	0.156	0.042	0.277	0.176
F20	-0.025	0.068	0.023	-0.098	-0.075	0.051
F22	-0.009	0.096	0.189	0.025	-0.007	-0.040
F23	0.171	0.221	0.149	0.045	-0.031	0.237
M10	0.193	0.350	0.150	0.276	0.051	0.140
M11	0.095	-0.036	0.164	0.271	0.272	0.253
M12	0.262	0.148	0.011	0.030	0.078	0.073
M13	0.422	0.412	0.474	0.638	0.122	0.167
M14	-0.030	0.190	0.122	0.443	0.148	-0.053
M15	0.182	0.287	0.211	0.363	0.245	0.252
M16	0.122	0.002	0.076	0.108	0.497	0.483
M17	-0.036	0.262	0.173	0.317	0.147	0.172
M18	-0.002	0.159	0.041	0.094	0.078	0.159
M22	-0.070	-0.036	-0.092	0.047	0.112	0.125
mean	0.068	0.013	0.011	0.162	0.122	0.136

Table 9: Correlation between RT and RD for successive 10-minute periods

3.1.6 Evaluation of Tonic Physiological Measures

3.1.6.1 Cardiovascular Measures

Cardiovascular data was collected with standard EKG electrode placements as well as with a Laser-Doppler Vibrometer (LDV). We collected cardiac data using the LDV by aiming the laser at the carotid artery on one side of the subject's neck. Data was recorded continuously during EPVT performance for all 20 subjects. The LDV used in this study was an early version of the LDV hardware and software system in development by our university research partners. It did not allow for automatic tracking of the subject. However, since subjects sat relatively still during the task, only occasional adjustments to the aim of the laser beam were necessary. The majority of the data analysis was conducted on the EKG based data. We monitored inter-beat intervals (IBI) through task performance.

Two types of analyses were conducted. The first dealt with the use of heart period data to evaluate ToT effects. This evaluated what we refer to as tonic changes associated with performing over time. The second analysis dealt with evaluating the relationship between heart period and performance of the running memory task. We argued that the running memory task involves anticipation of the event requiring a right mouse response. When the first odd integer appears, subjects are alerted, commit that information to memory and anticipate an odd integer on the next trial. The occurrence of the second odd integer acts to further alert the subject and cause the subject to anticipate the next stimulus. Our focus here was on heart period changes associated with "anticipation". There is a lengthy literature dealing with heart period changes associated with the "fixed foreperiod effect". The procedure is to require subjects to respond as rapidly as possible to a stimulus. That stimulus is preceded by a warning signal with the interval between warning and response requirement constant. Under these conditions one sees nice cardiac deceleration during the delay period, a response that does not decrease in amplitude over time, i.e. it is not an orienting response. We argued that the deceleration seen is in anticipation of an "imperative" stimulus. We thus expected that in our experiment subjects should demonstrate such cardiac deceleration (increase in heart period) in anticipation of the third odd integer. They should demonstrate such deceleration on both "hit" as well as "false alarm" trials. Whether the cardiac system would demonstrate such effects for "missed" trials was a question of interest. From personal experience and reports of subjects, some misses occur because the operator is uncertain about making a response.

3.1.6.2 Heart Period and ToT Effects

Heart period (HP) was averaged for successive 5-minute periods. We evaluated both changes in average heart period duration over successive 5-minute periods as well as changes in variability, reflected in the standard deviation (SD) of the mean heart period measure.

Figure 17 shows a graph of the average HP for all 20 subjects across time. For the first 30 minutes of task performance, the graph suggests a decrease in heart period (increase in rate) with an increase in heart period over the last 30 minutes so that heart period at time of termination of the experiment is roughly the same as it was at time of task initiation.

Heart rate variability appears to demonstrate a clean relationship with ToT. As is shown in Figure 18, the standard deviation around the mean for successive 5-minute periods increases in a systematic manner. We have an excellent linear fit with a regression line given by:

$$y = 2.307x + 53.893 \text{ and } R^2 = 0.944$$

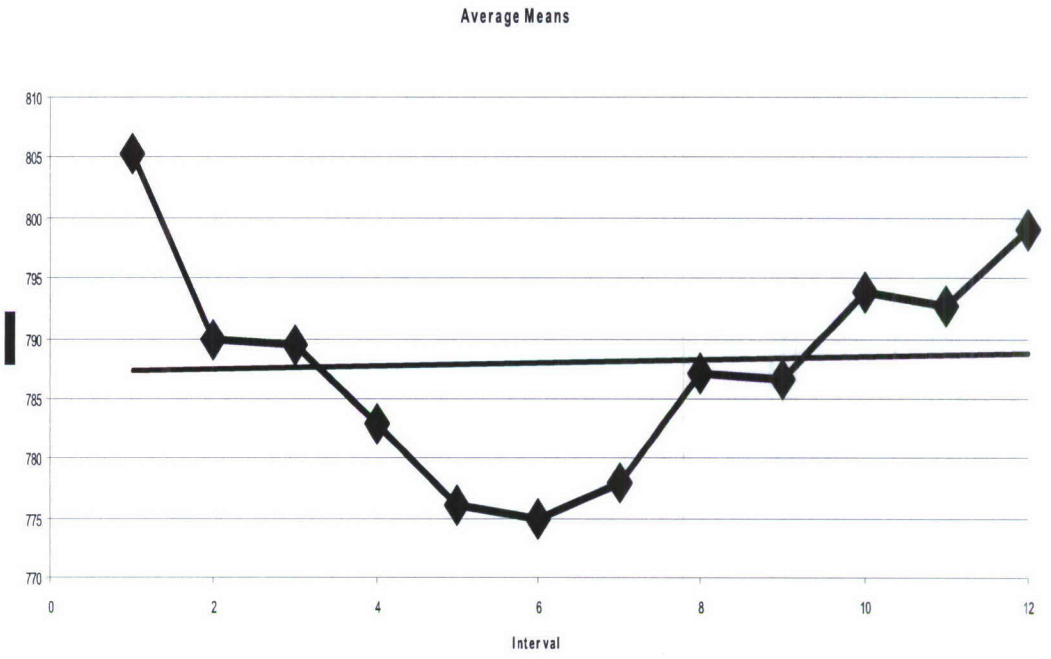


Figure 17: Heart Period Averaged Across Subjects

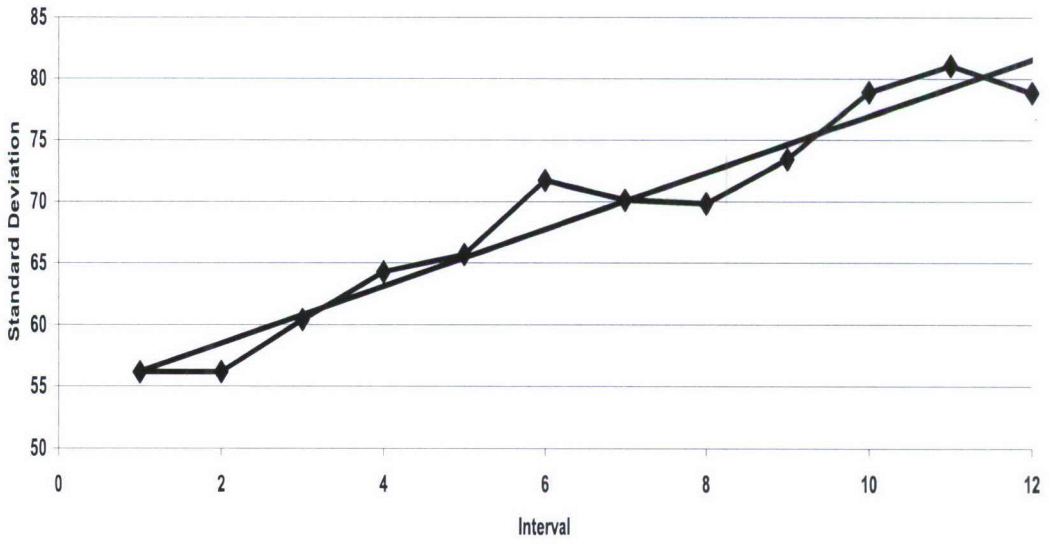


Figure 18: Heart Rate Variability

Nineteen of the twenty subjects exhibit the trend of increasing heart rate variability across time. The sole exception is subject F22. Using multi-level modeling, the average correlation across the 20 subjects is 0.607. This is statistically significant, and the significance test of the correlation coefficient is unbiased. This indicates that, on average, heart rate variability increases as ToT increases.

3.1.6.3 Pupil Diameter and ToT Effects

We wondered whether the increase in heart rate variability might be related to changes we had seen in pupil diameter as well as pupil diameter variability. These results are briefly presented below.

To gather information on how the pupil diameter changes across time, pupil diameter data was extracted from the data every 17 seconds. If the pupil diameter information was not available at the selected time point (generally caused by an eye blink during which the eye tracking system cannot locate the pupil), we selected a point 85 milliseconds after the eye tracking camera re-captured the pupil. At this point the pupil diameter information provided by the eye tracking system is again relatively stable. Data was then averaged over successive 5-minute periods, creating 12 time intervals.

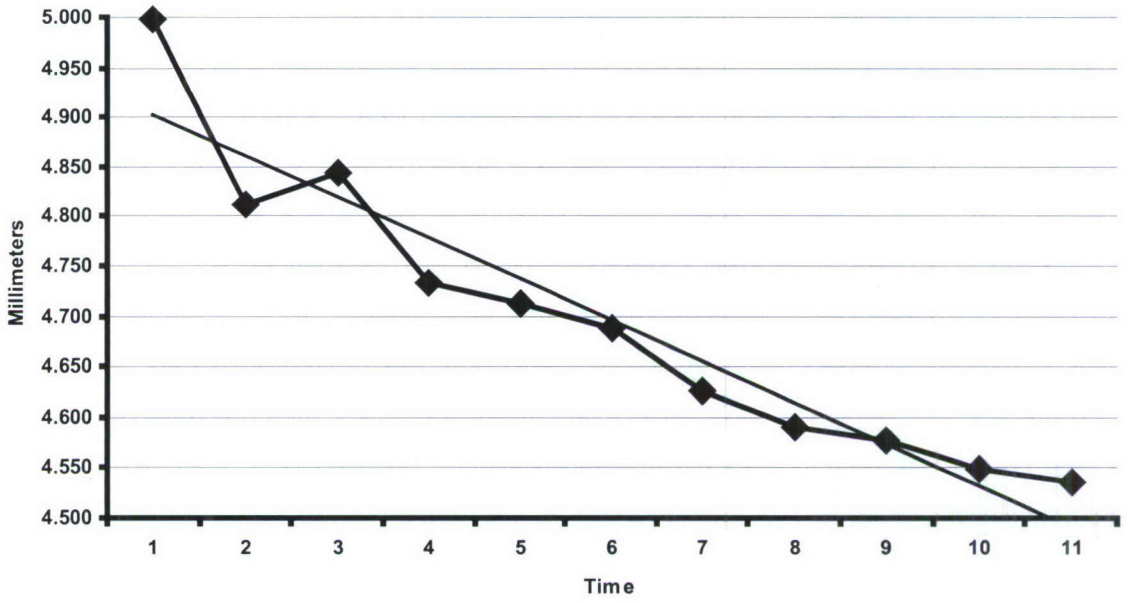


Figure 19 shows these results averaged over all 20 subjects.

The data is well represented by a simple linear equation:

$$y = -0.041x + 4.945 \text{ and } R^2 = 0.910.$$

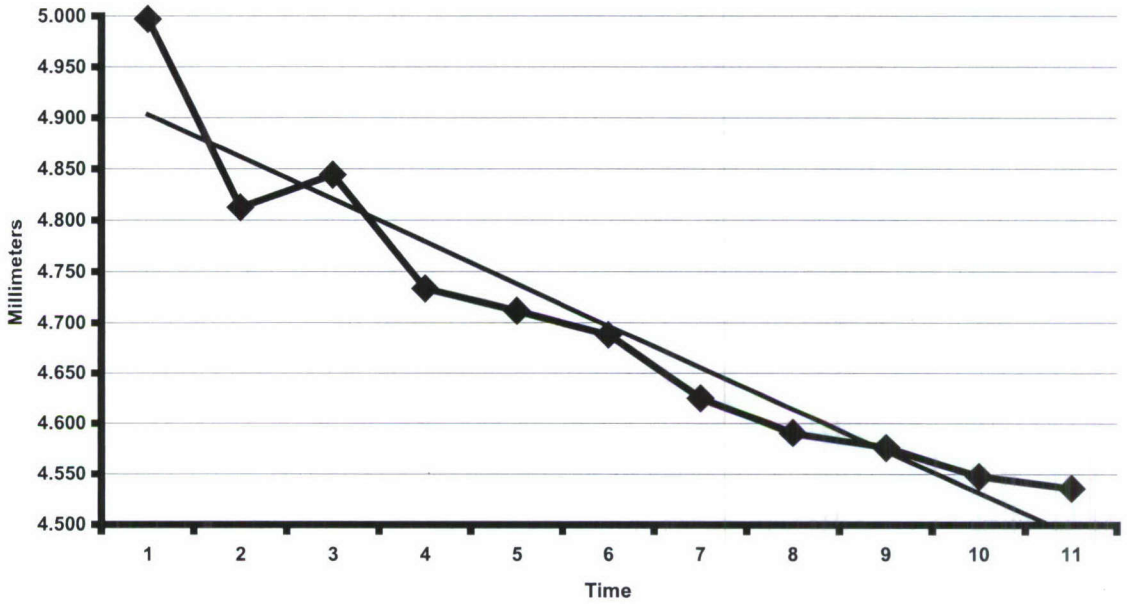


Figure 19: All Subjects Pupil Diameter Mean

Similar to the analysis done for the heart period measure, we analyzed pupil diameter variability as a function of ToT. We assessed the standard deviation of the mean values of pupil diameter for the same successive 5-minute periods used in the pupil diameter mean analysis. Figure 20 shows the results of this analysis.

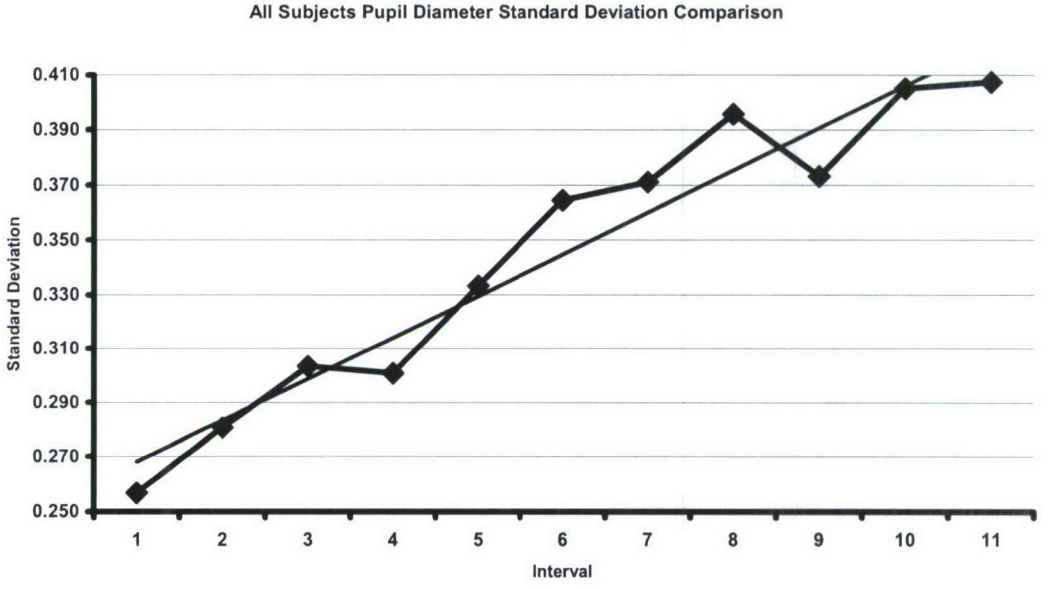


Figure 20: All Subjects Pupil Diameter Standard Deviation

From the graph shown in Figure 20, it appears that the effect (the increase in pupil diameter variability over time) is roughly equivalent to the amplitude of the effect of mean pupil increase over time. There is a good linear fit of a regression line to the pupil diameter variability data with a linear equation of:

$$y = 0.015x + 0.253 \text{ and } R^2 = 0.936.$$

This graph suggests that as ToT increases, there is a corresponding increase in pupil diameter variability.

Since we observed linear trends for both heart rate variability and pupil diameter variability, we wondered whether these were correlated events and that one could predict one from the other. Using our statistical modeling technique (two-level modeling), there is no evidence based on the chi-square test that the correlation between heart rate variability and pupil diameter variability differs significantly from zero. This suggests that these two measures may be associated with separate processes.

Random Effect	Standard Deviation	Variance Component	df	Chi-square	P-Value
ZPD slope, R1	0.153	0.023	19	26.283	0.122
level-1, E	0.814	0.662			

Table 10

3.1.6.4 Analysis of Overshoot Saccades

What is an overshoot saccade? It is a saccade that closely follows a preceding saccade, is in a direction opposite to that of the preceding saccade and is significantly smaller (in amplitude) than the first saccade. Such overshoot saccades have been observed by others, and the findings generally indicate that these saccades are more likely to occur in the abducting than the adducting eye. They are also reported to occur more frequently in "fatigued" subjects. We thus evaluated occurrence of such saccades in our 20 subjects. Since the camera system in use at the time only captured information from the right eye, we were not able to test the hypothesis concerning overshoot and the abducting eye. However, we observed that overshoot saccades occurred both to gaze shifts to the left and right. We thus included all overshoot saccades in our analysis.

For our analysis, we defined overshoot saccades as the second saccade in a pair of saccade where the second saccade:

1. occurs within 102 milliseconds of the termination of the first saccade
2. is in the opposite direction of the first saccade
3. is significantly smaller than the first saccade.

In our sample of 20 subjects, there appears to be a clear difference between those individuals who demonstrated many such events and those demonstrating relatively few such events. We have, in Figure 21 and Figure 22, displayed the overshoot events for these two groups separately. The subjects demonstrating a high level of overshoot saccades (Figure 22) were all male. Whether this gender difference will hold up in a larger sample is yet to be determined.

In the group of 15 subjects demonstrating low levels of such events, the average frequency per 5-minute period ranged from approximately 7 to 28 such events. As is shown in Figure 21, there appears to be steady increases in overshoot events for the initial 30 minutes of task performance with random fluctuations around the 30-minute level for the remainder of the experiment.

The five subjects for which data is shown in Figure 22 start out with much higher levels of overshoot saccades. Notice that the average number of overshoot saccades for the "normal" group is 7 in the initial 5-minute period, whereas the average number of overshoot saccades for this group in the same 5-minute period is 42. This group shows an increase in frequency of overshoot saccades for the initial 40 minutes of task performance followed by a reduction of such events for the last 20 minutes. The frequency of overshoot events during this last 20 minutes is still considerably above the starting level.

We can thus conclude that both groups demonstrate a ToT effect, and that there are marked individual differences in the frequency of occurrence of overshoot saccades across subjects. There is suggestive evidence that male subjects are more likely than female to demonstrate such overshoot.

Since there are suggestions in the literature that overshoot saccades as more likely to occur in the abducting than the adducting eye we reviewed occurrence of overshoot saccades to the left and right of the right eye. With the camera system available at the time this data was collected we were only able to sample one eye and consistently sampled the right eye. We find that the

majority (more than 60%) of overshoot saccades occur when gaze shift to the right is required. Thus the abducting eye has a higher frequency of such events. We see little in this data to suggest that there are ToT effects with respect to occurrence of this type of gaze shift.

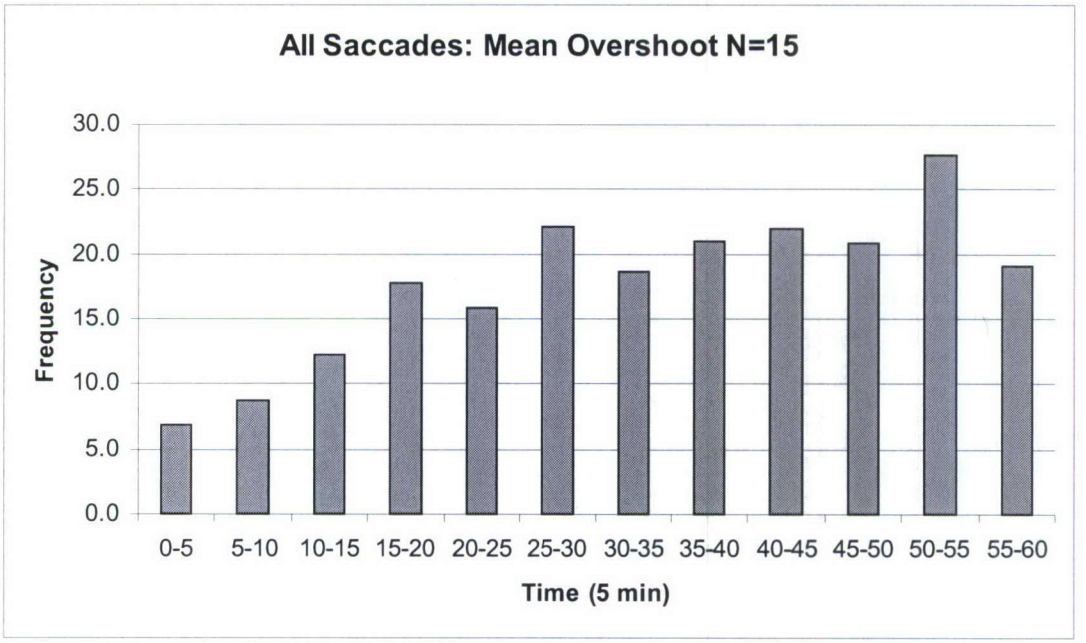


Figure 21: Mean Overshoot Frequency for 15 Subjects with Relatively Few Overshoot Saccades

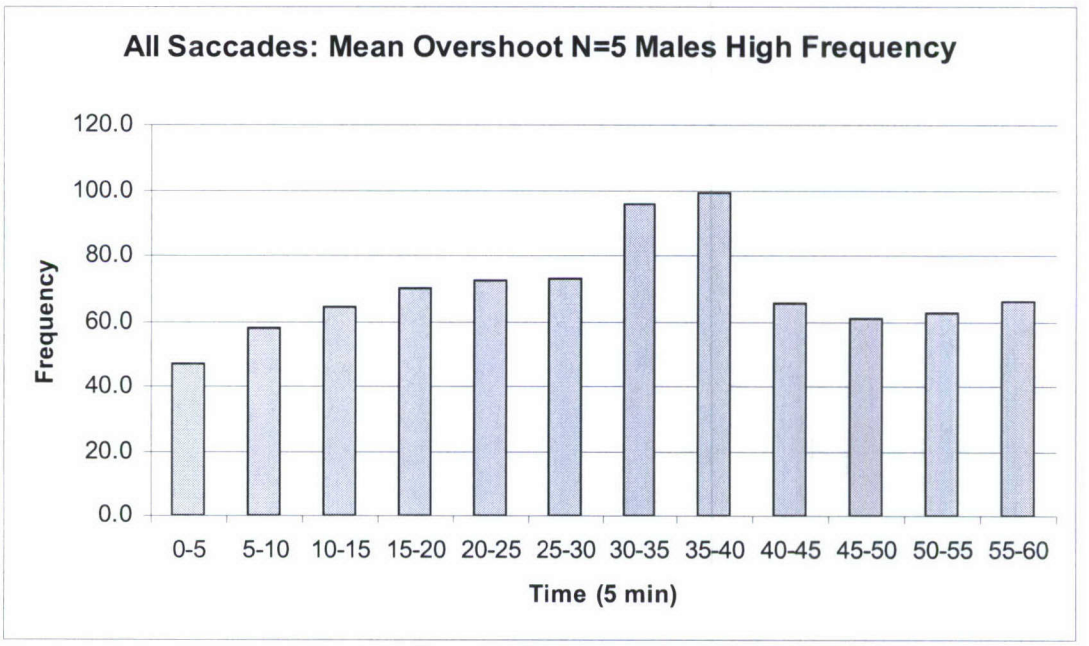


Figure 22: Mean Overshoot Frequency for 5 Subjects with a High Frequency of Overshoots

3.1.7 Evaluation of Phasic Physiological Measures

3.1.7.1 Using Pupillometric Measures to Detect Lapses in Alertness

The results of our analysis into using pupillometric measures to detect momentary lapses in alertness (phasic changes in alertness levels) are described in a paper that we have submitted for publication. That paper is included in this report as Appendix A.

3.1.7.2 Heart Period Response to Running Memory Component of the EPVT

This analysis dealt with the relationship between heart period and performance of the running memory task. We reasoned that the running memory task involves anticipation of the event requiring a right mouse response. When the first odd integer appears subjects are alerted, commit that information to memory and anticipate an odd integer on the next trial. The presentation of a second odd integer should be further alerting them to expect the third odd integer, requiring a right mouse click response on their part. We evaluated heart period for five inter-beat intervals (IBIs) preceding and five IBIs following enactment of a right mouse press response. We could do this analysis when a response was made, regardless of whether it was a correct response or a false alarm. For trials on which they missed making a running memory task response, we identified, for the individual, the point in time when a response would most likely have been made, based on the timing of correct responses. Adequate data for conducting these analyses were available for only five of the 20 subjects.

Rather than using absolute heart period we used the first IBI identified as the covariate and expressed subsequent intervals as a ratio of that interval. Thus a positive ratio is associated with an increase in the IBI (decrease in heart rate). We then plotted the average across the five subjects for the figures below.

The enactment of a correct response (hit) is associated with an increase in heart period up to and immediately following response enactment with a rapid decrease for the last 4 IBIs. Figure 23 depicts these results.

The pattern of heart period change associated with enactment of a false alarm closely mirrors the above pattern to the point of response enactment and the IBI immediately following responding. However, following that event the IBI remains elevated for the remaining intervals sampled. We have no idea whether subjects were aware of making a response when no response was required. Cardiac activity, however, appeared to make that distinction. See Figure 24 below.

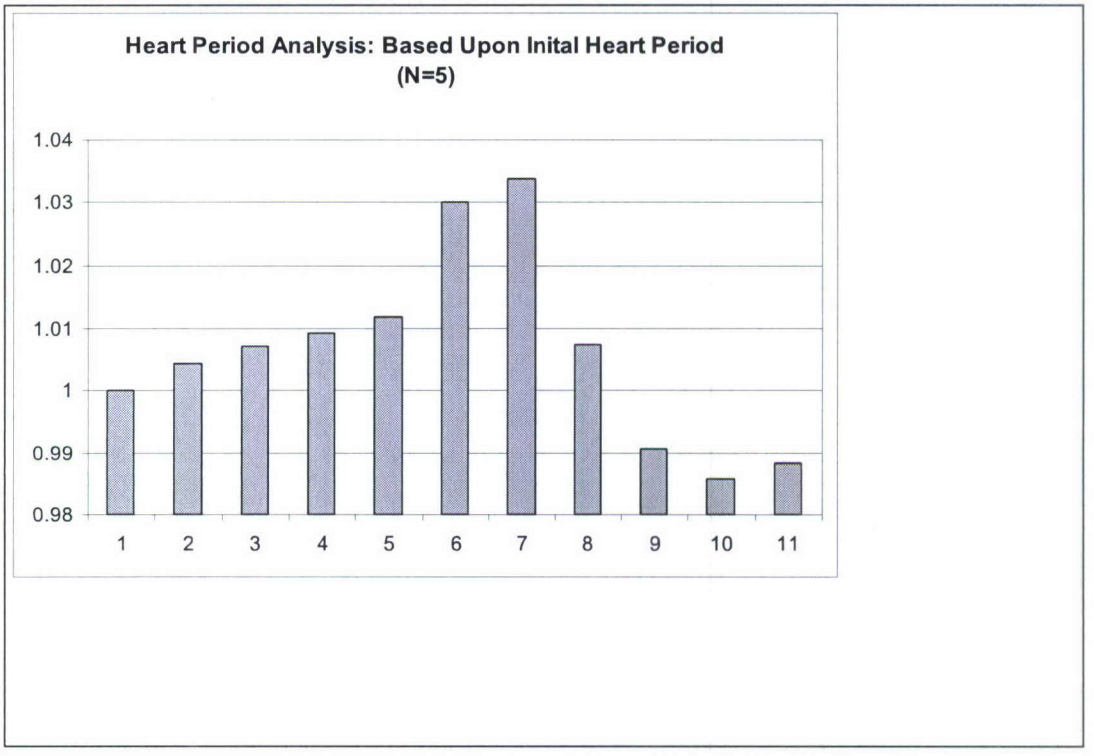


Figure 23

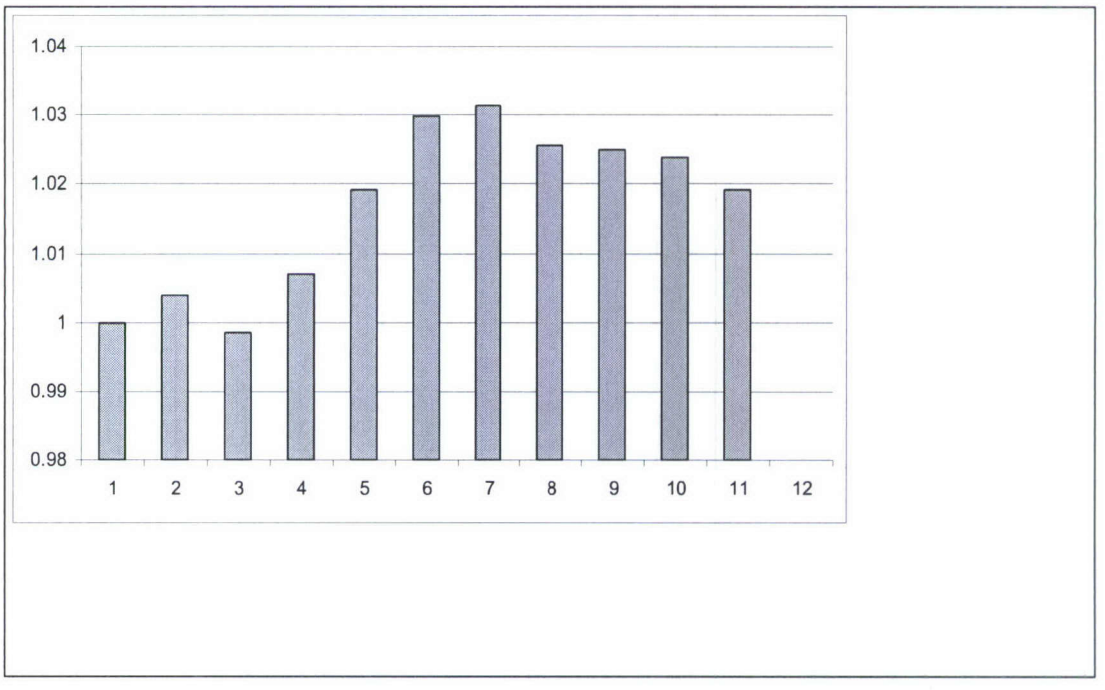


Figure 24

Last but not least we have the pattern associated with missed signals. Even here we see that whatever controls the heart, it appears to be responding to the occurrence of a sequence of three odd integers, though not to the extent seen for either correct responses or false alarms. Peak IBI occurs at the appropriate time and, as was true for the false alarms the IBI remains elevated for the following 5 IBIs. See Figure 25 below.

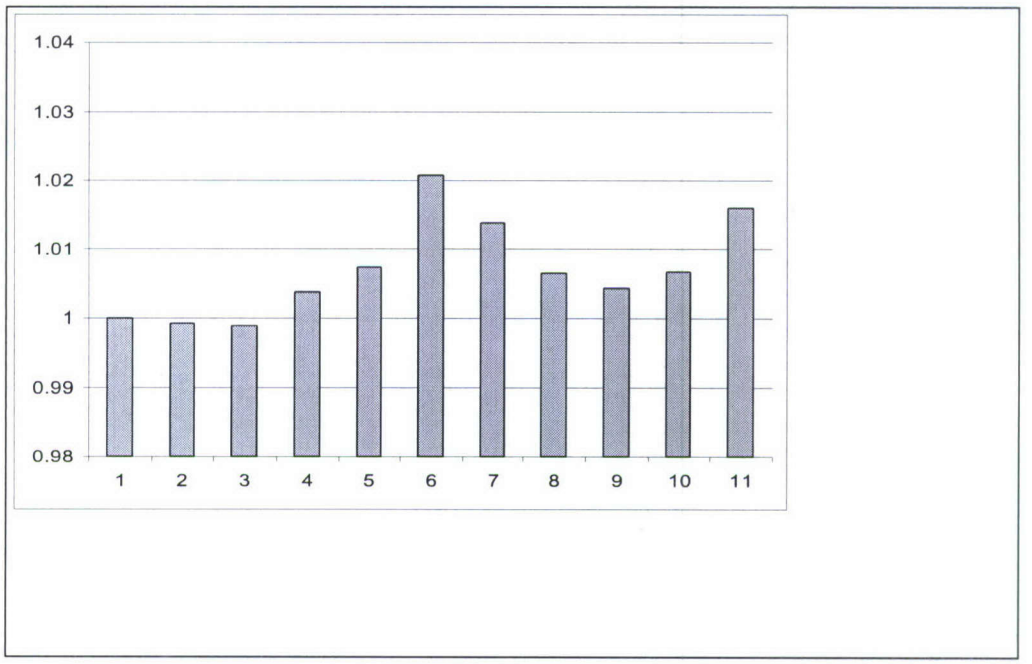


Figure 25

These results showing the heart period response to the running memory component of the EPVT are intriguing and need replication.

3.1.7.3 Timing of Eye Blinks: Random or Cognition Related?

The results of our analysis into the timing of eye blinks are described in a separate paper. That paper is included in this report as Appendix B.

3.2 Evaluation of Cognitive Complexity's Effect on Reaction Time (RT)

The results of our analysis into cognitive complexity's effect on reaction time are also described in a separate paper. That paper is included in this report as Appendix C.

3.3 Evaluation of EPVT Performance of Sleep Disturbed Individuals

In collaboration with the Washington University Sleep Disorder Center, we evaluated EPVT performance for a group of 10 HIV positive patients, all of whom complained of sleep disturbance. These subjects were considerably older than our sample of 20 college students whose data was presented earlier. Because of time limitations we required this patient population to perform the task for 30 minutes. These patients had significantly more difficulty in understanding instructions concerning the running memory component of the task. Difficulties ranged from not knowing the difference between an odd and even number to understanding instructions as reflected in performance during the 5-minute training period and then not responding during the 30 minutes of task performance. They were well motivated but demonstrated difficulty in performing the running memory task. Analysis was thus limited to RT for the detection component of the task. Figure 26 depicts mean reaction time for the 30-minute task. In the figure and those that follow the patient population is labeled BEPVT. We have also included results for the initial 30 minutes of task performance for our college age group. That population is labeled EPVT.

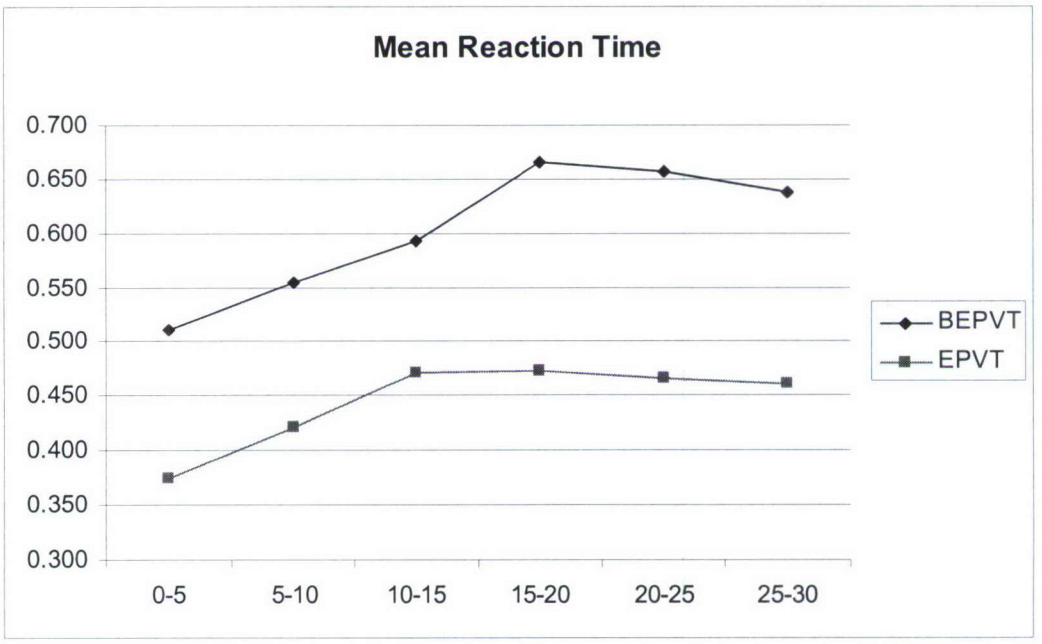


Figure 26: Mean Reaction Time Comparison

It is readily apparent that RT is significantly slower in the patient group and that they reach an asymptotic level of performance somewhat later than the college age subjects.

We evaluated the 25 fastest and 25 slowest responses made during the 30-minute period for both groups. These results are depicted in Figure 27 and Figure 28. For 25 fastest responses (Figure 27), we see that for the patient sample 9.3 of the 25 fastest responses occur in the initial 5-minute period (37%). For the student group, 7.1 of the 25 fastest responses (28%) occur in those first 5

minutes. The pattern of decreasing incidence of fast responses over time appears to evolve somewhat more rapidly for the patient (BEPVT) group. The pattern for slowest responses (Figure 28) appears to increase steadily in both groups over the 30-minute period.

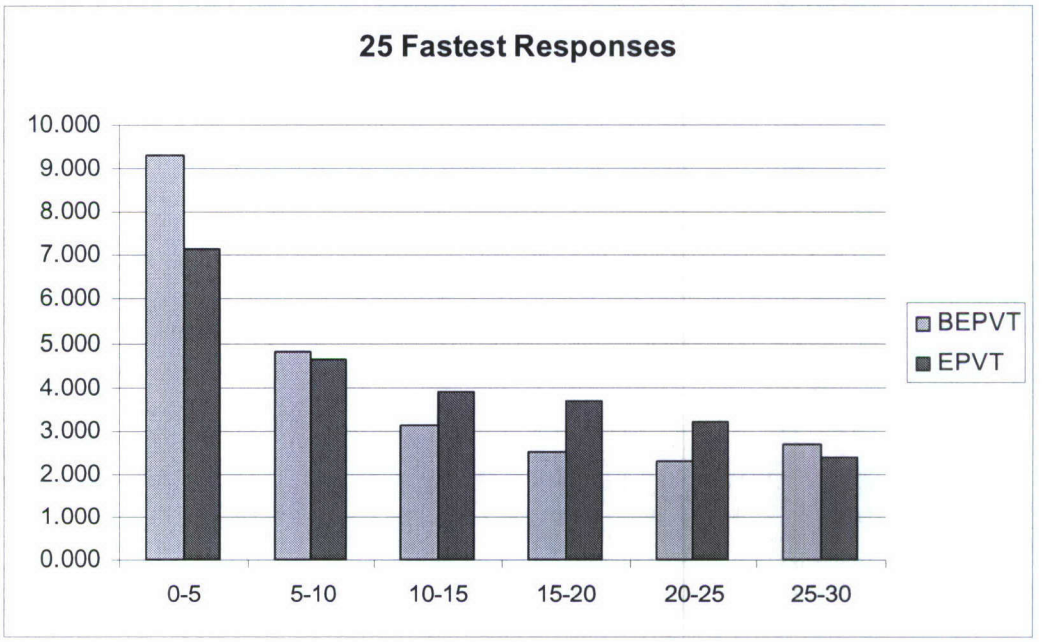


Figure 27

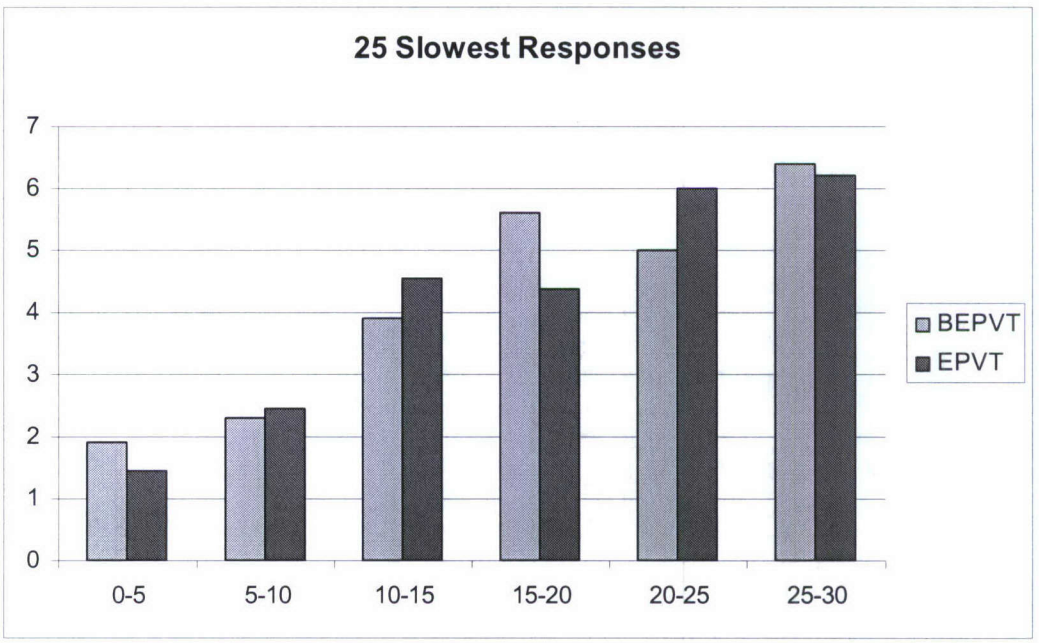


Figure 28

Since we had oculometric measures available for both groups, we also evaluated the time between stimulus presentation and the first saccade shifting gaze to the target location. The average delay for the two groups was 0.364 seconds and 0.381 seconds respectively for HIV positive and the student group. Thus, surprisingly, the delay in responding seen in the HIV+ group is not attributable to any slowing of perceptual processes, i.e., the detection of peripherally presented events. The slowing is most likely associated with the minimal information processing required in making the decision to press the left mouse button following detection of an incrementing count-up timer.

3.4 Evaluation of Repeat Reliability of EPVT Performance

To determine whether individuals respond in a consistent fashion, both behaviorally as well as physiologically, we studied a small sample of subjects who performed the EPVT on three or four occasions. Subjects A, E, and J performed the task on 4 occasions. Subject F1 performed the task on 3 occasions. These subjects received the feedback on the running memory task that is described above.

3.4.1 RT as a function of ToT

The performance of the subjects in this study was markedly superior to the students who participated in the study described in section 2.1. Figure 29 presents the average RT within successive 5-minute periods for each subject.

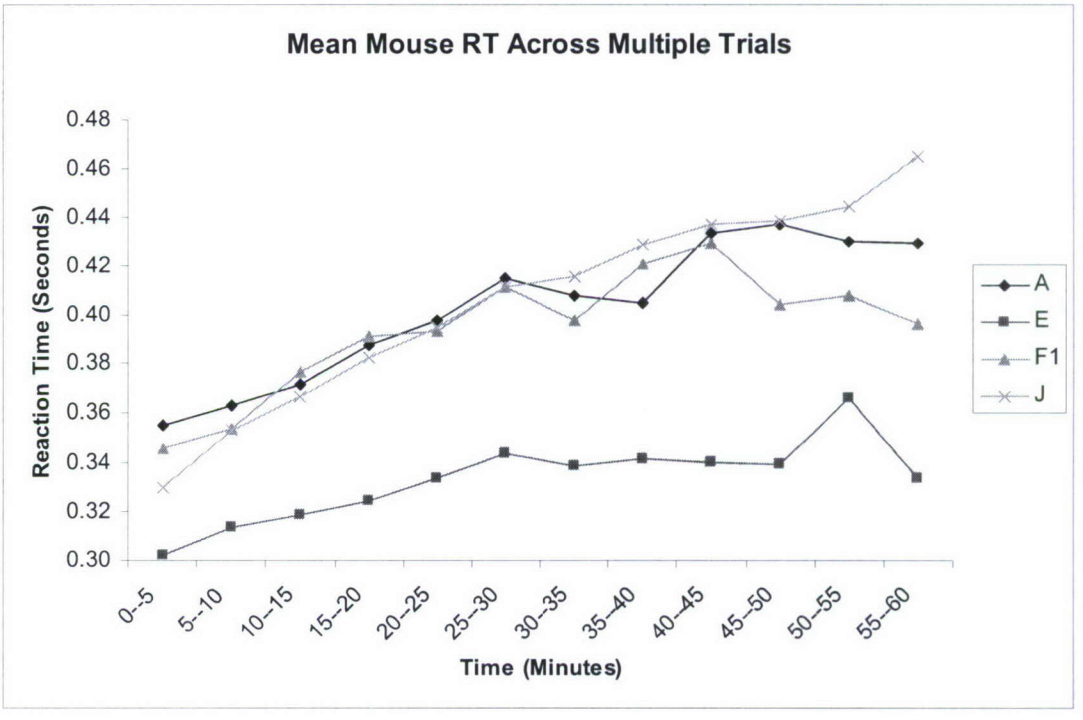


Figure 29

Three of the subjects (A, F1, and J) demonstrated similar patterns of performance. Subject E not only responded more rapidly than the others during the initial 5 minutes of task performance, but also demonstrated less of a decline over the 60 minutes of task performance.

3.4.2 Response duration measures

Average duration of the left mouse button press is presented in Figure 30. Three of the four subjects demonstrate an increase in average response duration with subject A maintaining a stable level.

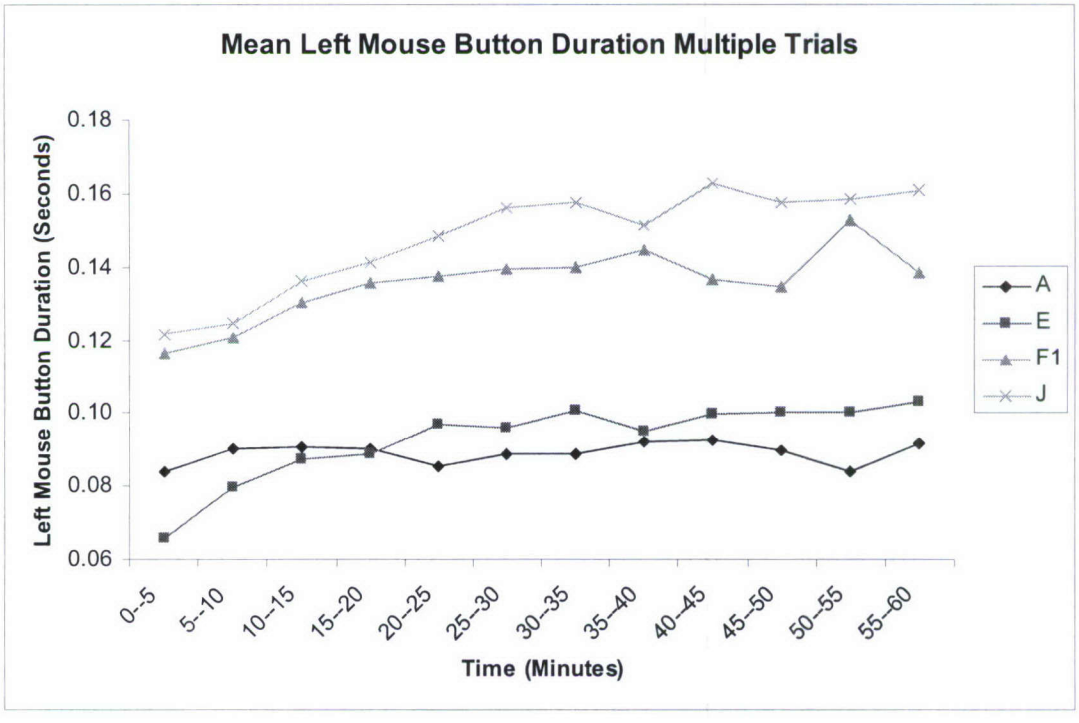


Figure 30

3.4.3 Long latency responses

We evaluated the distribution of the 50 longest latency responses. Similar to the findings for our study of 20 college age subjects, there is a significant increase in long latency responses as a function of ToT. Figure 31 depicts the average number of long latency responses for each subject. The following 4 figures (Figure 32, Figure 33, Figure 34, and Figure 35) depict the results for each of the three or four trials for each subject.

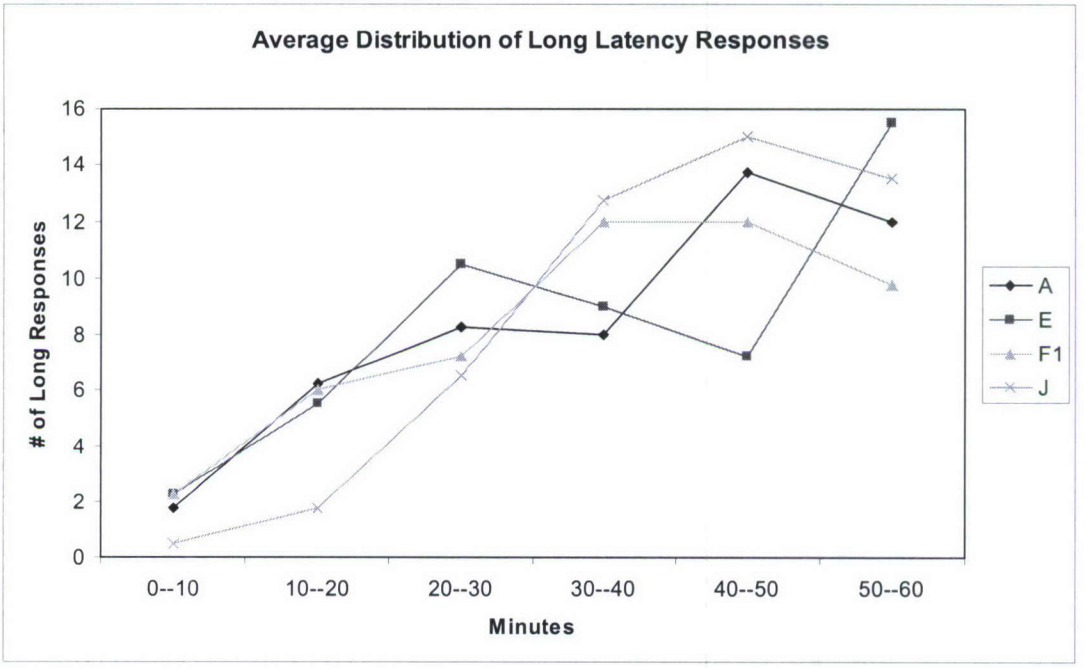


Figure 31

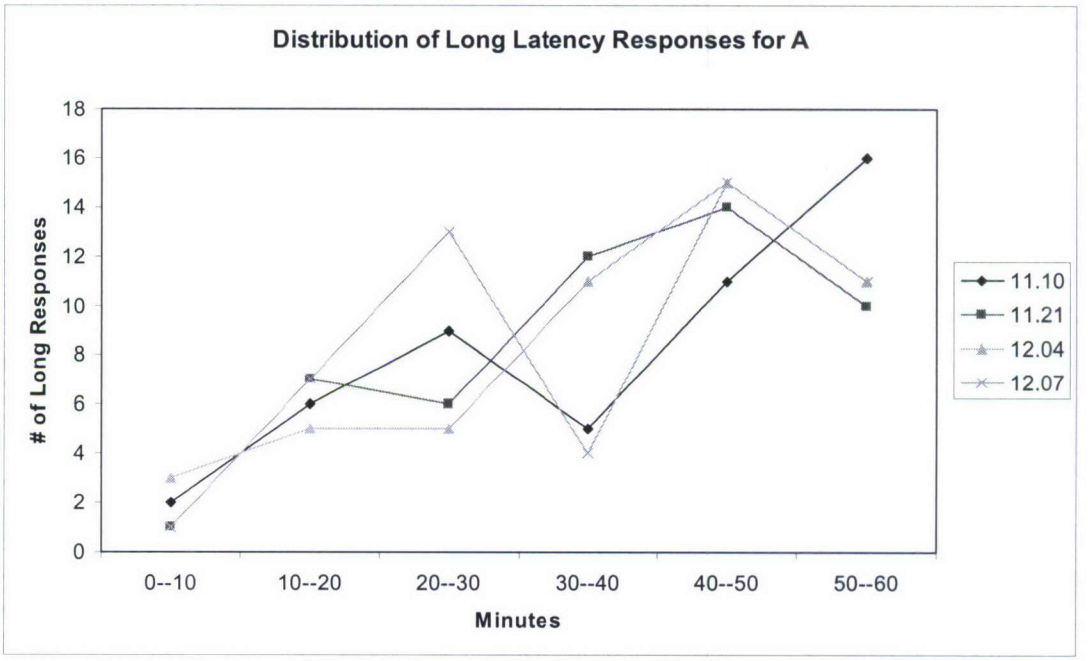


Figure 32

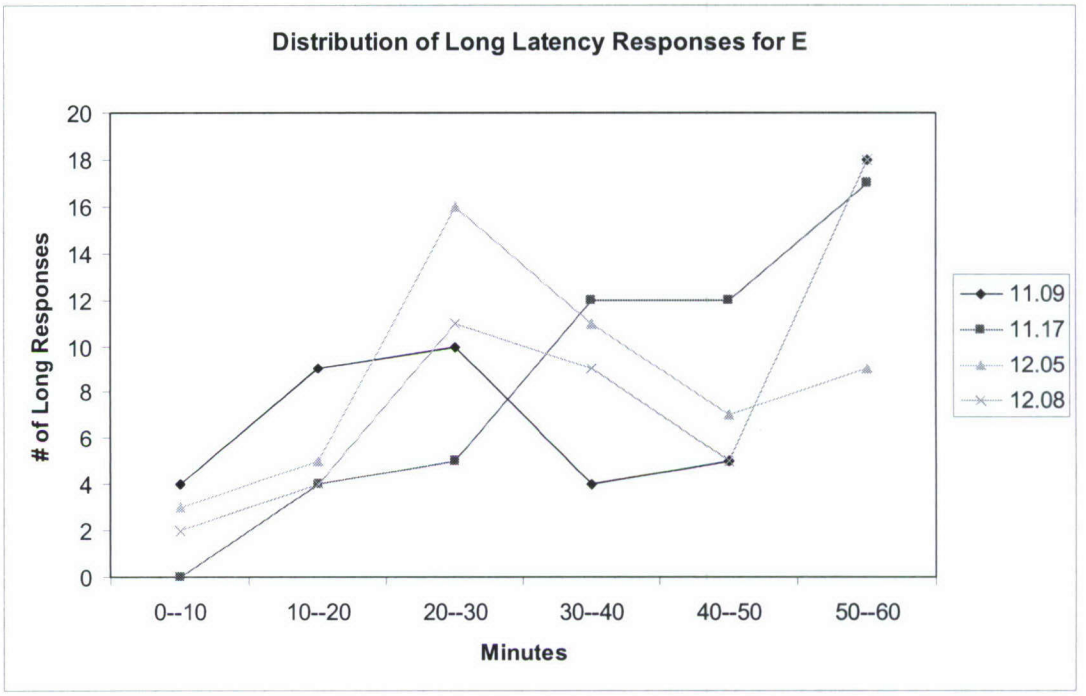


Figure 33

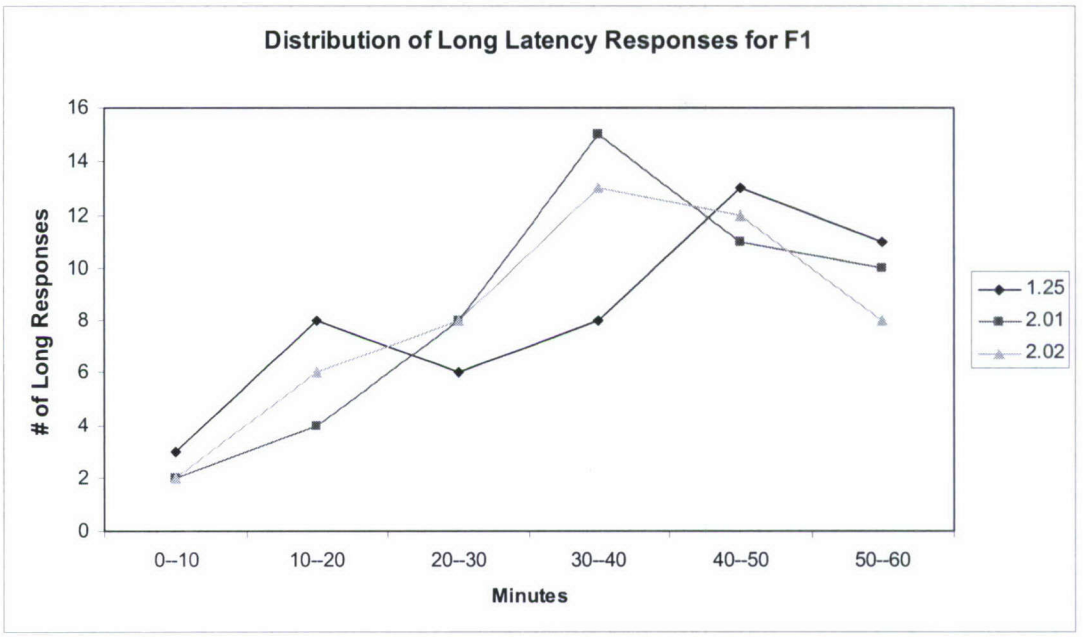


Figure 34

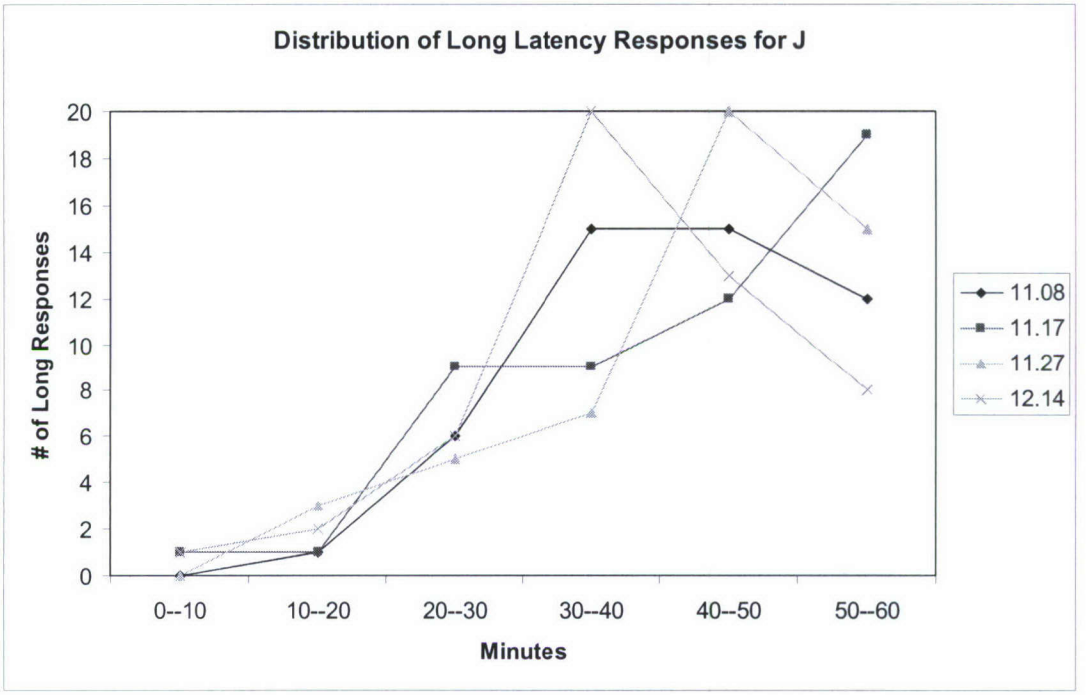


Figure 35

Though there are some day-to-day variations in the distribution of these events, the general pattern appears to be quite similar across days. Subject E is the only one who shows a decline in long latency events between minutes 30 and 50 for three of four experiments. Subjects appear to have unique patterns of such decrements in performance. Subject J, for example, differs from the other three in showing the fewest such events during the initial 10 minutes of task performance.

3.4.4 Saccadic eye movements as reflectors of alertness loss

Can saccadic eye movements be used to identify lapses in alertness? There is a literature dealing with the reduction in saccade velocity as a function of fatigue. Most of this research has used technologies that sample the eye move frequently than is true of video technology. The latter technology, as exemplified in our equipment, samples the eye at 60 Hz. We believe that this is too slow to allow for the measurement of peak saccade velocity. It does allow for the identification of saccade duration to an accuracy of 17 milliseconds.

Saccades can be initiated rapidly or slowly; the slope of the eye movement can be slow or fast; and the saccade can terminate slowly or rapidly. Saccades with slow onset or termination are identified as glissades and are associated with "fatigue" (Bahill and Stark, 1975). Wang (1998) studied glissadic saccades as a possible measure of vigilance. He found that when information acquisition is important, the likelihood of a glissadic saccade is small. He reports that, "for blink-free saccades, the occurrence of glissadic saccades increased with time-on-task for anticipatory and return (from target) saccades but not for stimulus elicited saccades"

For the current study, we identified long duration saccades in the following manner:

a. We correlated saccade duration with saccade with saccade amplitude for successive five minutes of task performance.

b. We used the regression equation developed for the initial five minutes of task performance as a predictor of the duration for each saccade in subsequent periods based on the amplitude of the saccade. Thus, for any saccade in the subsequent 5-minute periods, we plug the saccade's amplitude into the regression equation to arrive at a predicted duration for that saccade.

c. We used this predicted duration to identify saccades that were "outliers" in the direction of being of longer duration than predicted.

d. We identified such outliers in successive five or ten minute periods.

e. We determined, by inspection, whether the saccade met our criteria of being an outlier because of saccade velocity or glissadic activity. Then we discriminated between onset and terminal glissades.

f. Since the number of saccades in a period differed not only between individuals but also between successive periods, we expressed our metric in terms of a ratio where the denominator was the total number of saccades identified for the period.

3.4.4.1 Correlation between saccade amplitude and duration

We correlated the relationship between saccade duration and amplitude expecting correlations not as large as those seen when one correlates saccade amplitude with peak velocity, but still in a range acceptable for allowing us to perform the above operations. Table 11 through Table 14 show the correlations for successive five-minute periods of task performance for every subject for each occasion on which they performed the EPVT.

The correlations, which are labeled ρ in the tables, are "reasonable" especially that for the initial 5-minute period. The exception to this is subject E. We suspect that the low correlations for this subject are attributable to the fact that he demonstrated many more overshoot saccades than was true of the other subjects. The return component of an overshoot saccade is more variable in duration and amplitude than is true for other saccades of the same amplitude. It is this variability that we believe accounts for the low correlations for this subject.

Subject A									
11.10.06	**Right Eye**		**Left Eye**		**11.21.06**	**Right Eye**		**Left Eye**	
Time	**N**	**ρ**	**N**	**ρ**	**Time**	**N**	**ρ**	**N**	**ρ**
00-05	101	0.646	105	0.712	00-05	108	0.715	100	0.683
05-10	96	0.595	102	0.731	05-10	128	0.645	129	0.669
10-15	77	0.642	79	0.706	10-15	113	0.536	122	0.756
15-20	110	0.722	111	0.713	15-20	122	0.666	127	0.599
20-25	120	0.602	122	0.690	20-25	120	0.593	133	0.644
25-30	113	0.610	122	0.693	25-30	141	0.627	142	0.685
30-35	99	0.596	88	0.681	30-35	113	0.565	111	0.536
35-40	128	0.570	127	0.689	35-40	135	0.712	147	0.710
40-45	145	0.603	139	0.667	40-45	160	0.526	168	0.670
45-50	166	0.578	179	0.680	45-50	152	0.586	150	0.646
50-55	175	0.473	190	0.669	50-55	147	0.603	146	0.668
55-60	170	0.553	169	0.598	55-60	114	0.493	120	0.596
12.04.06	**Right Eye**		**Left Eye**		**12.07.06**	**Right Eye**		**Left Eye**	
Time	**N**	**ρ**	**N**	**ρ**	**Time**	**N**	**ρ**	**N**	**ρ**
00-05	118	0.460	109	0.643	00-05	139	0.642	145	0.609
05-10	117	0.499	119	0.618	05-10	135	0.694	140	0.658
10-15	111	0.654	113	0.627	10-15	146	0.660	148	0.741
15-20	109	0.492	112	0.555	15-20	136	0.688	138	0.606
20-25	130	0.529	140	0.552	20-25	141	0.574	145	0.623
25-30	138	0.448	144	0.510	25-30	151	0.510	166	0.690
30-35	107	0.578	108	0.619	30-35	103	0.470	107	0.614
35-40	117	0.311	117	0.387	35-40	117	0.550	132	0.541
40-45	142	0.389	137	0.540	40-45	150	0.648	153	0.679
45-50	142	0.283	141	0.347	45-50	155	0.507	150	0.498
50-55	134	0.268	137	0.366	50-55	151	0.510	146	0.605
55-60	127	0.262	125	0.283	55-60	125	0.607	122	0.584

Table 11: Saccade Amplitude/Duration Correlations for Subject A

Subject E									
11.09.06	Right Eye		Left Eye		11.17.06	Right Eye		Left Eye	
Time	N	ρ	N	ρ	Time	N	ρ	N	ρ
00-05	90	0.491	91	0.372	00-05	101	0.292	102	0.526
05-10	106	0.321	108	0.425	05-10	101	0.273	103	0.112
10-15	105	0.304	102	0.114	10-15	115	0.307	112	0.415
15-20	83	0.367	86	0.379	15-20	118	0.432	112	0.412
20-25	112	0.396	115	0.208	20-25	108	0.305	101	0.275
25-30	110	0.297	108	0.093	25-30	124	0.258	128	0.378
30-35	86	0.130	82	0.303	30-35	110	0.141	111	0.163
35-40	124	0.358	128	0.148	35-40	109	0.359	115	0.248
40-45	95	0.307	94	0.277	40-45	111	0.273	109	0.190
45-50	49	0.158	56	0.106	45-50	121	0.293	114	0.363
50-55	data loss		47	0.220	50-55	131	0.365	128	0.301
55-60	31	0.333	28	-0.058	55-60	81	0.186	86	0.354
12.05.06	Right Eye		Left Eye		12.05.06	Right Eye		Left Eye	
Time	N	ρ	N	ρ	Time	N	ρ	N	ρ
00-05	112	0.262	106	0.194	00-05	89	0.453	91	0.430
05-10	93	0.349	89	0.278	05-10	107	0.415	105	0.205
10-15	107	0.345	101	0.371	10-15	101	0.397	99	0.438
15-20	121	0.301	124	0.442	15-20	105	0.346	108	0.307
20-25	112	0.232	119	0.392	20-25	116	0.414	122	0.232
25-30	109	0.110	123	0.264	25-30	112	0.228	118	0.198
30-35	118	0.295	123	0.176	30-35	98	0.294	107	0.217
35-40	126	0.448	142	0.382	35-40	120	0.309	121	0.158
40-45	107	0.149	113	0.097	40-45	102	0.064	99	0.253
45-50	112	0.395	116	0.305	45-50	108	0.266	112	0.095
50-55	133	0.149	133	0.232	50-55	109	0.257	110	0.204
55-60	99	0.270	99	0.230	55-60	96	0.366	95	0.201

Table 12: Saccade Amplitude/Duration Correlations for Subject E

Subject F1										
1.25.07	Right Eye		Left Eye			2.1.07	Right Eye		Left Eye	
Time	N	ρ	N	ρ		Time	N	ρ	N	ρ
00-05	372	0.575	374	0.568		00-05	360	0.672	363	0.624
05-10	331	0.532	321	0.508		05-10	365	0.615	373	0.587
10-15	368	0.624	361	0.540		10-15	372	0.583	372	0.579
15-20	443	0.468	437	0.595		15-20	377	0.450	397	0.514
20-25	452	0.589	433	0.580		20-25	383	0.618	398	0.618
25-30	501	0.566	485	0.537		25-30	384	0.550	376	0.551
30-35	415	0.604	386	0.663		30-35	390	0.586	396	0.551
35-40	388	0.666	393	0.588		35-40	351	0.554	370	0.494
40-45	459	0.586	461	0.590		40-45	380	0.462	353	0.514
45-50	384	0.592	369	0.595		45-50	373	0.495	358	0.338
50-55	461	0.642	458	0.639		50-55	405	0.494	402	0.523
55-60	458	0.559	456	0.602		55-60	361	0.591	355	0.620
2.2.07	Right Eye		Left Eye							
Time	N	ρ	N	ρ						
00-05	318	0.437	324	0.550						
05-10	356	0.517	363	0.451						
10-15	361	0.593	358	0.516						
15-20	389	0.456	385	0.522						
20-25	338	0.495	335	0.446						
25-30	366	0.556	348	0.439						
30-35	337	0.491	325	0.376						
35-40	311	0.376	298	0.340						
40-45	321	0.420	332	0.452						
45-50	329	0.459	320	0.435						
50-55	379	0.485	367	0.465						
55-60	341	0.541	335	0.462						

Table 13: Saccade Amplitude/Duration Correlations for Subject F1

Subject J									
11.8.06	**Right Eye**		**Left Eye**		**11.17.06**	**Right Eye**		**Left Eye**	
Time	**N**	**ρ**	**N**	**ρ**	**Time**	**N**	**ρ**	**N**	**ρ**
00-05	126	0.685	134	0.782	00-05	114	0.711	108	0.620
05-10	134	0.613	139	0.723	05-10	116	0.574	120	0.641
10-15	142	0.614	143	0.654	10-15	116	0.551	119	0.695
15-20	139	0.600	140	0.662	15-20	123	0.591	129	0.606
20-25	171	0.482	176	0.499	20-25	104	0.623	104	0.649
25-30	162	0.448	161	0.417	25-30	115	0.602	116	0.621
30-35	144	0.544	146	0.438	30-35	83	0.495	85	0.526
35-40	100	0.201	102	0.327	35-40	119	0.524	124	0.567
40-45	89	0.507	85	0.232	40-45	116	0.402	116	0.517
45-50	150	0.468	144	0.377	45-50	139	0.433	145	0.444
50-55	184	0.342	182	0.390	50-55	120	0.312	119	0.241
55-60	116	0.414	104	0.363	55-60	141	0.573	141	0.442
11.27.06	**Right Eye**		**Left Eye**		**12.14.06**	**Right Eye**		**Left Eye**	
Time	**N**	**ρ**	**N**	**ρ**	**Time**	**N**	**ρ**	**N**	**ρ**
00-05	105	0.628	104	0.810	00-05	127	0.598	130	0.721
05-10	123	0.689	122	0.771	05-10	100	0.464	100	0.746
10-15	113	0.606	115	0.759	10-15	99	0.675	95	0.765
15-20	141	0.548	148	0.647	15-20	114	0.682	117	0.597
20-25	116	0.469	118	0.670	20-25	125	0.579	124	0.414
25-30	141	0.544	136	0.484	25-30	91	0.381	88	0.325
30-35	119	0.563	114	0.556	30-35	126	0.516	123	0.479
35-40	124	0.472	127	0.400	35-40	57	0.580	60	0.564
40-45	113	0.632	110	0.558	40-45	116	0.352	123	0.513
45-50	128	0.347	126	0.300	45-50	130	0.473	127	0.436
50-55	138	0.361	127	0.457	50-55	128	0.513	122	0.494
55-60	139	0.367	137	0.368	55-60	147	0.552	141	0.490

Table 14: Saccade Amplitude/Duration Correlations for Subject J

3.4.4.2 Saccade frequency as a function of ToT

We next determined whether there were changes in saccade frequency as a function of ToT as well as evaluating within subject consistency in saccade frequency and between subject differences on this measure. Table 15 through Table 18 present these results.

	Left Eye	Right Eye
Session 11.10		
00-10	207	197
10-20	190	187
20-30	244	233
30-40	215	227
40-50	318	311
50-60	359	345
SUM	1533	1500
Session 11.21		
00-10	229	236
10-20	249	235
20-30	275	261
30-40	258	248
40-50	318	312
50-60	266	261
SUM	1595	1553
Session 12.04		
00-10	228	235
10-20	225	220
20-30	284	268
30-40	225	224
40-50	278	284
50-60	262	261
SUM	1502	1492
Session 12.07		
00-10	285	274
10-20	286	282
20-30	311	292
30-40	239	220
40-50	303	305
50-60	268	276
SUM	1692	1649

Table 15: Subject A Saccade Frequency

	Left Eye	Right Eye
Session 11.09		
00-10	199	196
10-20	188	188
20-30	223	222
30-40	210	210
40-50	150	144
50-60	75	31
SUM	1045	991
Session 11.17		
00-10	205	202
10-20	224	233
20-30	229	232
30-40	226	219
40-50	223	232
50-60	214	212
SUM	1321	1330
Session 12.05		
00-10	195	205
10-20	225	228
20-30	242	221
30-40	265	244
40-50	229	219
50-60	232	232
SUM	1388	1349
Session 12.08		
00-10	196	196
10-20	207	206
20-30	240	228
30-40	228	218
40-50	211	210
50-60	205	205
SUM	1287	1263

Table 16: Subject E Saccade Frequency

	Left Eye	Right Eye
Session 11.08		
00-10	273	260
10-20	283	281
20-30	337	333
30-40	292	262
40-50	288	239
50-60	248	300
SUM	1721	1675
Session 11.17		
00-10	228	230
10-20	248	239
20-30	220	219
30-40	209	202
40-50	261	255
50-60	260	261
SUM	1426	1406
Session 11.27		
00-10	226	228
10-20	263	254
20-30	254	257
30-40	241	243
40-50	236	241
50-60	264	277
SUM	1484	1500
Session 12.14		
00-10	230	227
10-20	212	213
20-30	212	216
30-40	183	183
40-50	250	246
50-60	263	275
SUM	1350	1360

Table 17: Subject J Saccade Frequency

	Left Eye	Right Eye
Session 11.25		
00-10	695	703
10-20	798	811
20-30	918	953
30-40	779	803
40-50	830	843
50-60	914	919
SUM	4934	5032
Session 02.01		
00-10	736	725
10-20	769	749
20-30	774	767
30-40	766	774
40-50	711	753
50-60	757	766
SUM	4513	4534
Session 02.02		
00-10	687	674
10-20	743	750
20-30	683	704
30-40	623	648
40-50	652	650
50-60	702	720
SUM	4090	4146

Table 18: Subject F1 Saccade Frequency

We see, in these tables that:

a. There are no consistent differences in saccade frequency as a function of ToT.

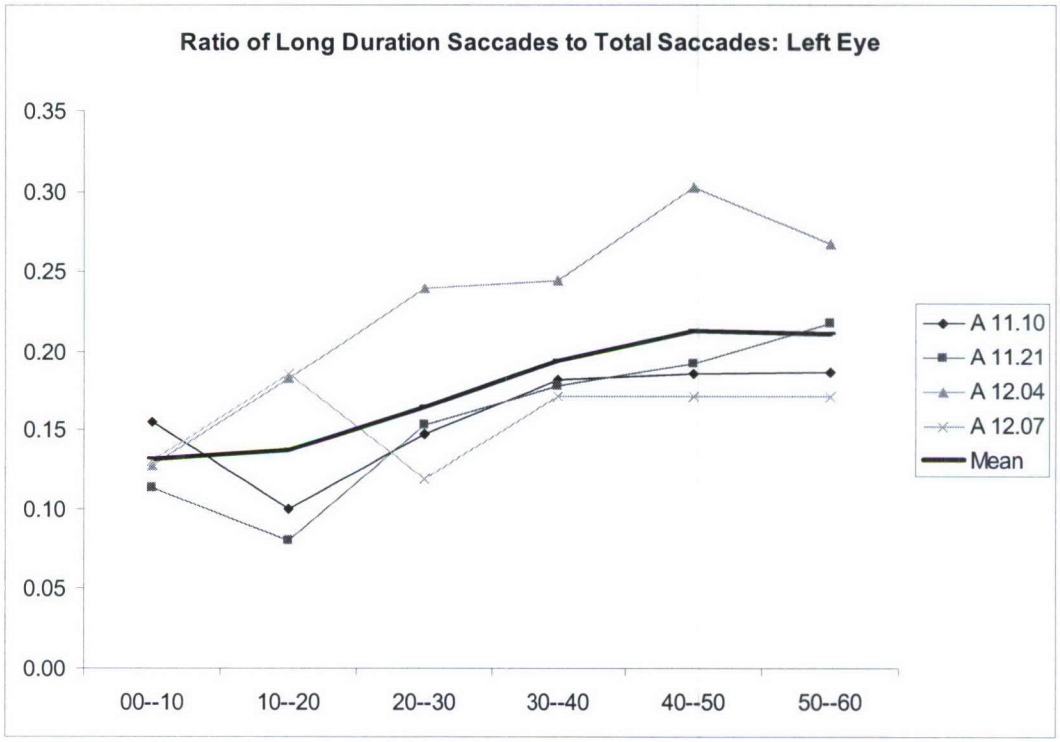
   That observation must be qualified by the fact that as a function of ToT, there was an increase in frequency of data loss. In this study, we did not attempt to control for this. This was especially true for Subject E, Session 11.09, where for the last 20 minutes data acquisition for both the left and right eye was "sporadic".

b. There are no consistent differences in the frequency with which our algorithm identified saccades from the left and right eye.

c. There are major differences in saccade frequency across subjects. In particular, subject F1 demonstrates saccade frequencies far in excess of any of the other three subjects.

3.4.4.3 Identification of long duration saccades

Because of the marked differences in saccade frequency across subjects, we decided to identify long duration saccades not in absolute terms but as a ratio of total number of saccades sampled during the period. The following figures present the results of these analyses.



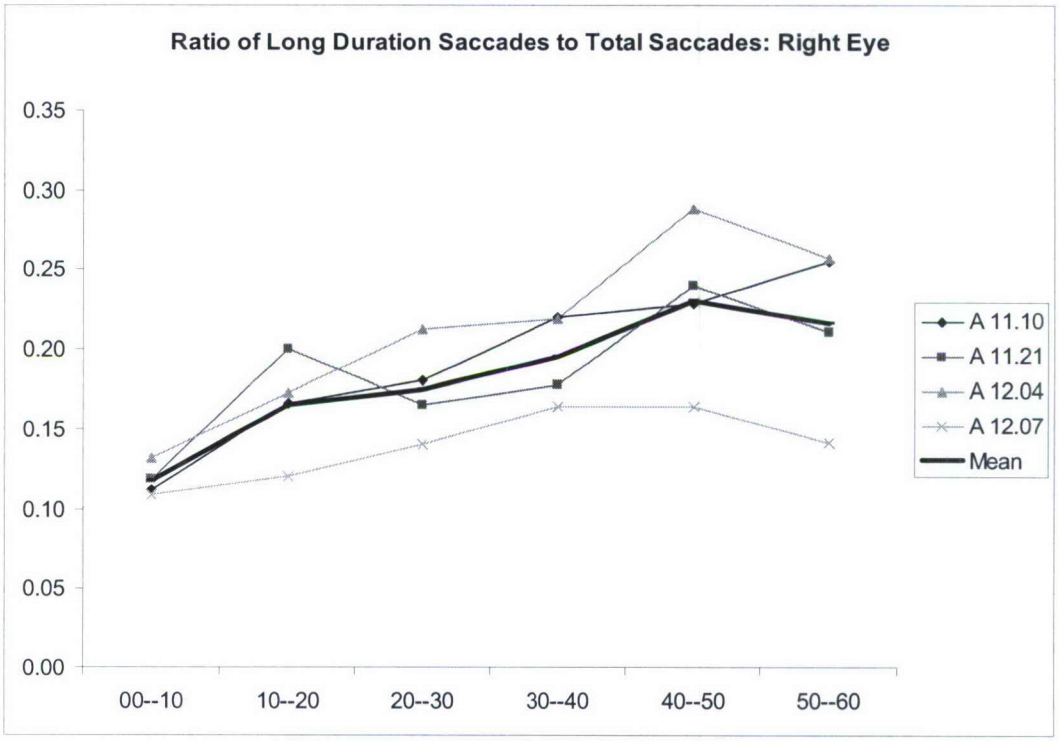
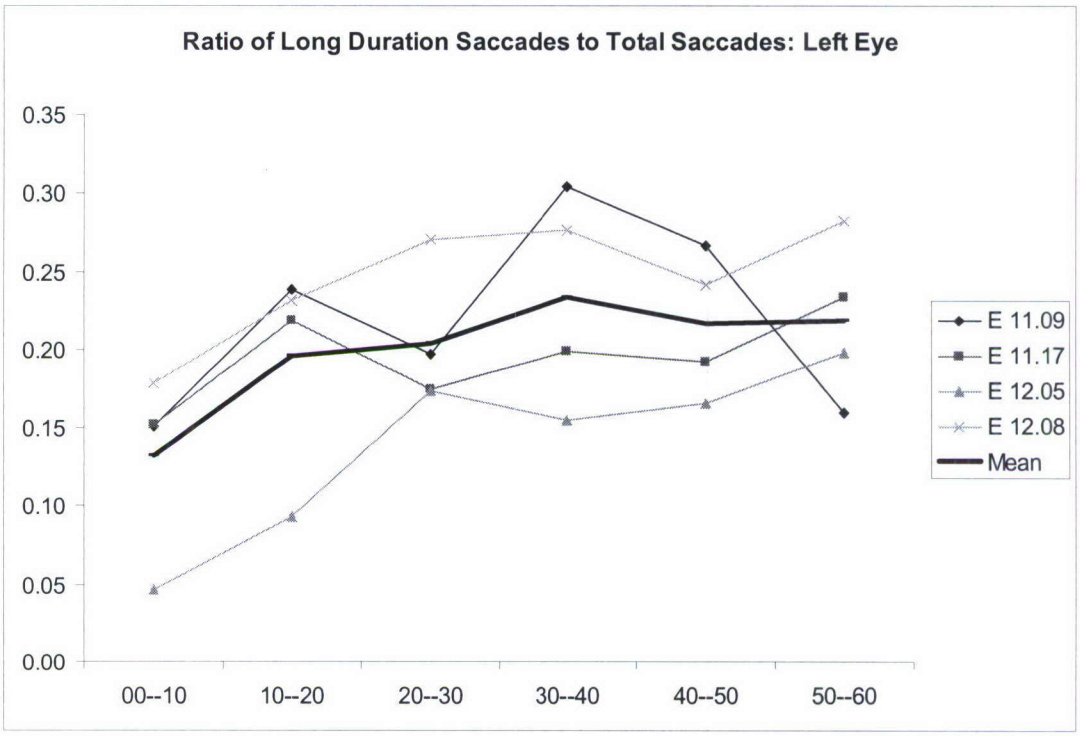


Figure 36: Ratios of Long Duration Saccades – Subject A



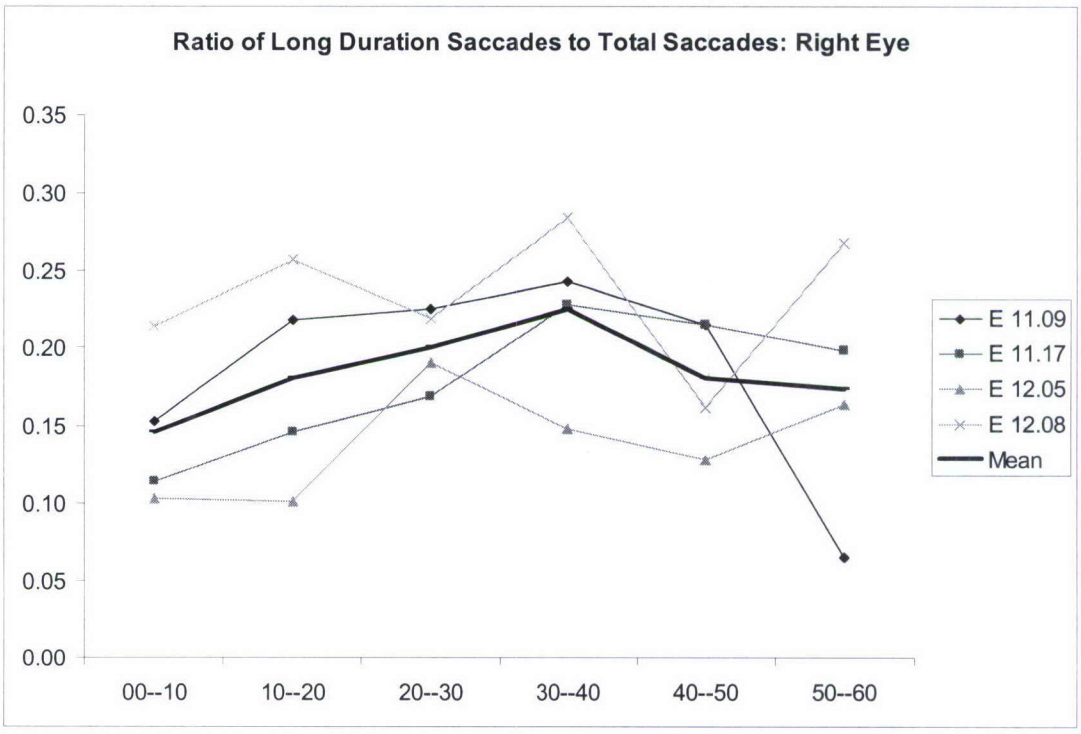


Figure 37: Ratios of Long Duration Saccades – Subject E

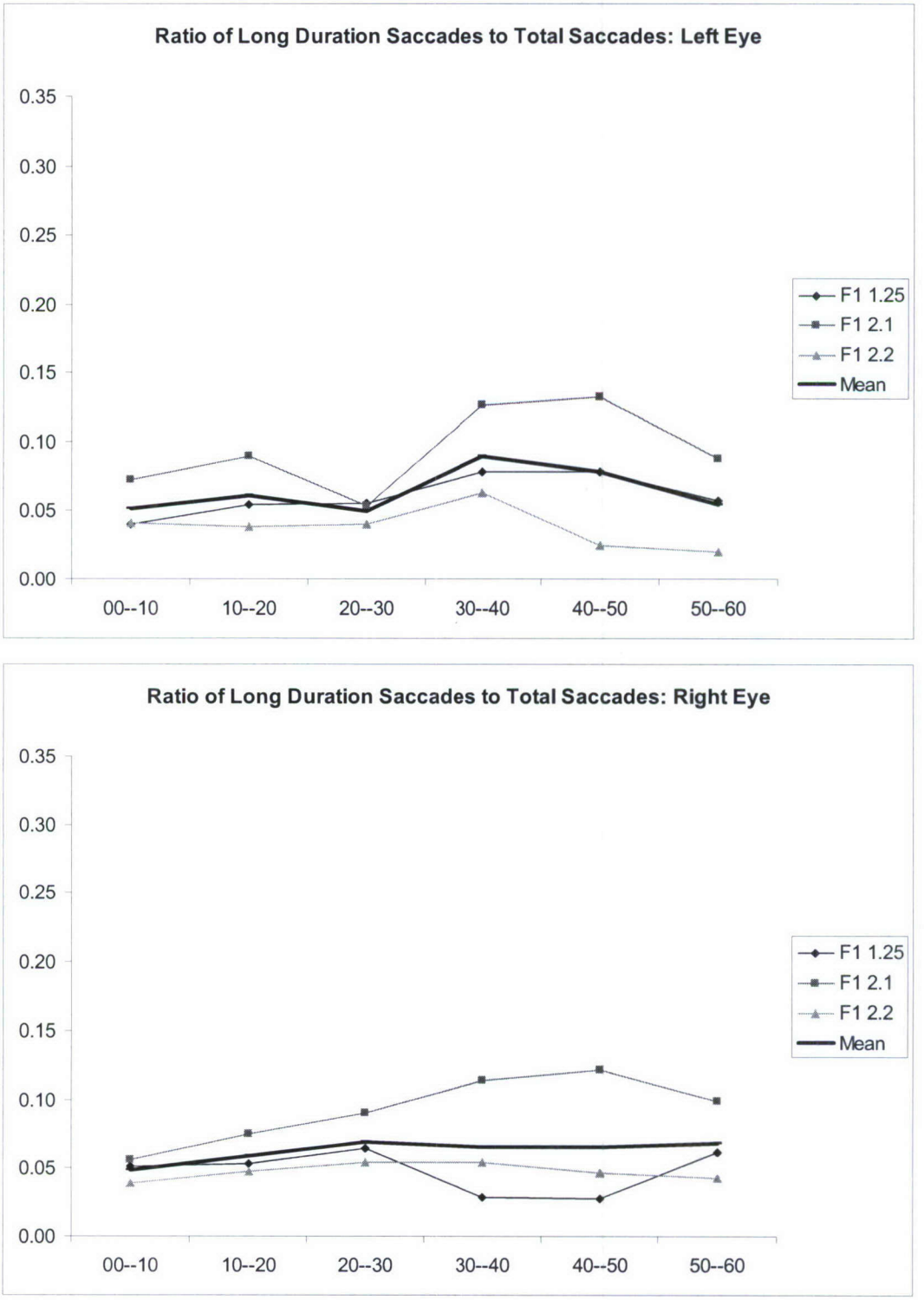
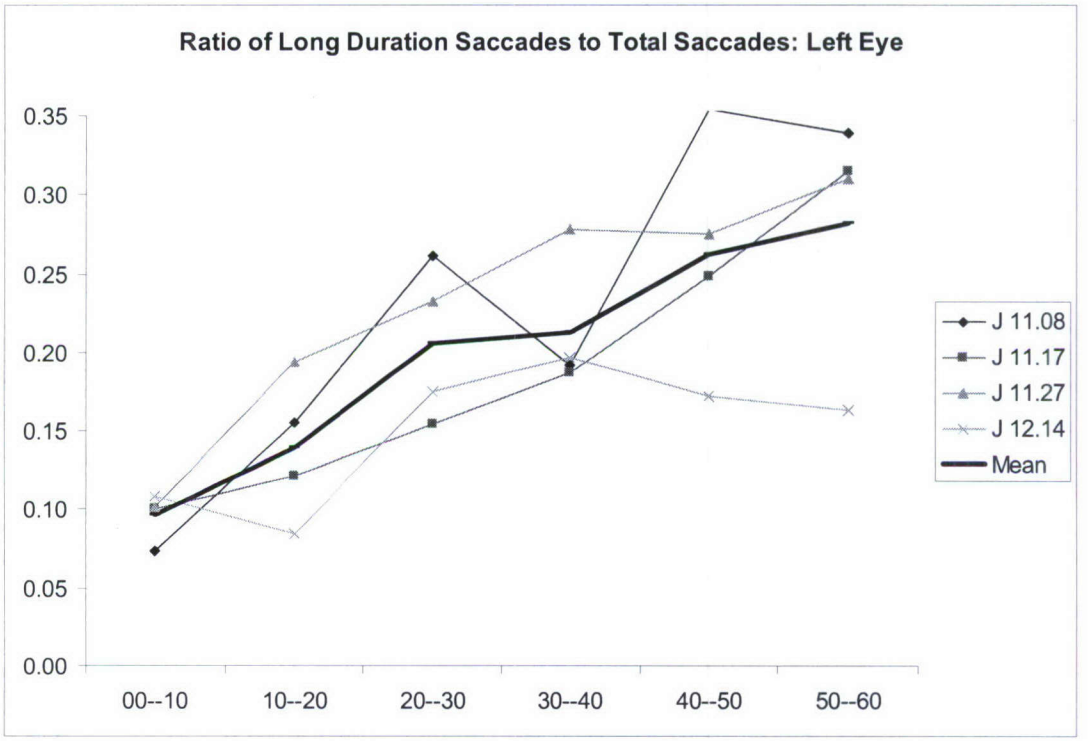


Figure 38: Ratios of Long Duration Saccades – Subject F1



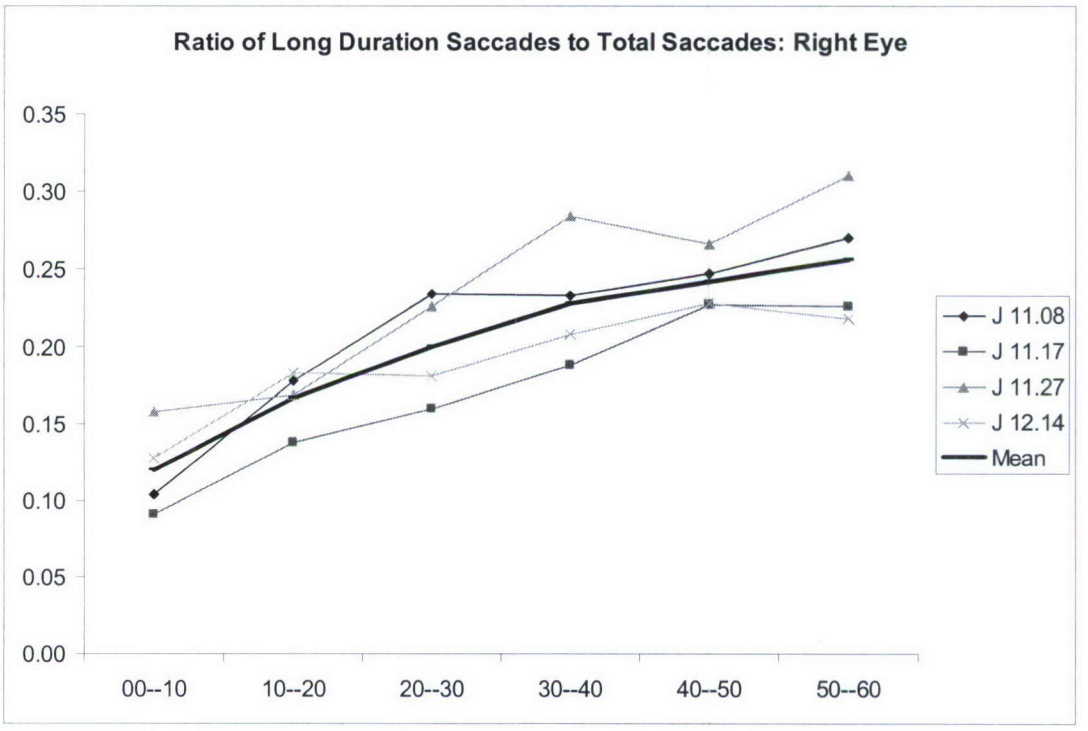


Figure 39: Ratios of Long Duration Saccades – Subject J

There are marked individual differences in the usefulness of this measure to demonstrate increases in such events as a function of ToT. Subjects A and J demonstrate consistent patterns of increasing frequency of long duration saccades as a function of ToT and demonstrate this effect on all four occasions. Subject F1 demonstrates increases for the initial 30 minutes of task performance and the pattern becomes more variable for the last 30 minutes. This is the subject with the largest number of saccades per unit time. Subject E does not demonstrate any consistent pattern. This is the subject for whom the correlation between saccade duration and amplitude was small. We suspect that the unreliability with which we identified long duration saccades for this individual was responsible for the lack of relationship identified.

3.4.4.4 Glissades or saccades, which component reflects alertness lapses?

We will not present all the data here but will summarize our results. We find that it is the slope component of the saccade that appears to be most responsive to ToT effects. Our analysis of both the initiating as well as terminating glissade suggests that they do not change much as a function of ToT. Past research (McGregor and Stern 1996), utilizing electrooculographic procedures to capture saccadic eye movements, demonstrated that saccade duration during a blink adapts to the duration of the blink. Not every saccade associated with a blink demonstrates this phenomenon, but enough do to demonstrate that the increase in saccade duration as a function of time is affected by blink duration. In the current study, utilizing camera based procedures to capture saccades, we cannot evaluate the latter during a blink. We do find, however, that as a function of ToT there is some dissociation between these two events, i.e., blinks are more likely to occur independent of saccades later in task performance and we can now identify and evaluate saccades returning gaze from the peripheral location back to a central location that earlier were not available due to blink occurrence. The increase in long duration saccades may thus be an artifact of being able to observe such saccades late in task performance while early in task performance they were obscured by the co-occurrence of a blink. Further analysis of our data, breaking the long duration saccades down by whether associated with a gaze shift toward or away from the target location should clear up this issue.

3.5 Evaluation of Head Movements in Acquisition of Information

The results of our evaluation of head movements in the acquisition of information are described in a separate paper. That paper is included in this report as Appendix D.

3.6 Demonstration of the Effects of Caffeine on EPVT Performance

As part of our effort to utilize a DoD organization to test the utility of the EPVT in identifying lapses in alertness and to test our software for real-time assessment and feedback of operator alertness status, we developed a collaborative relationship with Col. M. Russo at the Army Aeromedical Research Laboratories at Ft. Rucker Alabama. Our portion of a research effort labeled "Through the eyes of pilots" focused on the effect of sleep deprivation on EPVT performance. That study, unfortunately, was abandoned early this year without any data having been collected.

In February 2007, we delivered our equipment to the laboratory and after training the involved personnel to utilize our equipment we decided, on the last day of our visit, to involve them in a

mini-study. The PI for this STTR contract (Dr. John A. Stern) agreed to abstain from having his usual two cups of coffee in the morning. He was then tested with the EPVT for a 30 minute run. Next he ingested the equivalent of two cups of coffee by chewing two Army provided pieces of gum laced with caffeine and after half an hour of rest performed the EPVT for a second time. The experiment was a success, the local operator performed admirably in acquiring the data and the results were in line with expectations.

3.6.1 Results

Figure 40 depicts mean reaction times over the 30-minute period for the pre and post caffeine conditions. RT is consistently longer under the pre-caffeine condition. The increase in RT as a function of ToT does not discriminate between the two conditions, except that for the last 5 minutes of task performance under the post-caffeine condition there is a sudden and marked improvement in performance while the converse is true under the pre-caffeine condition.

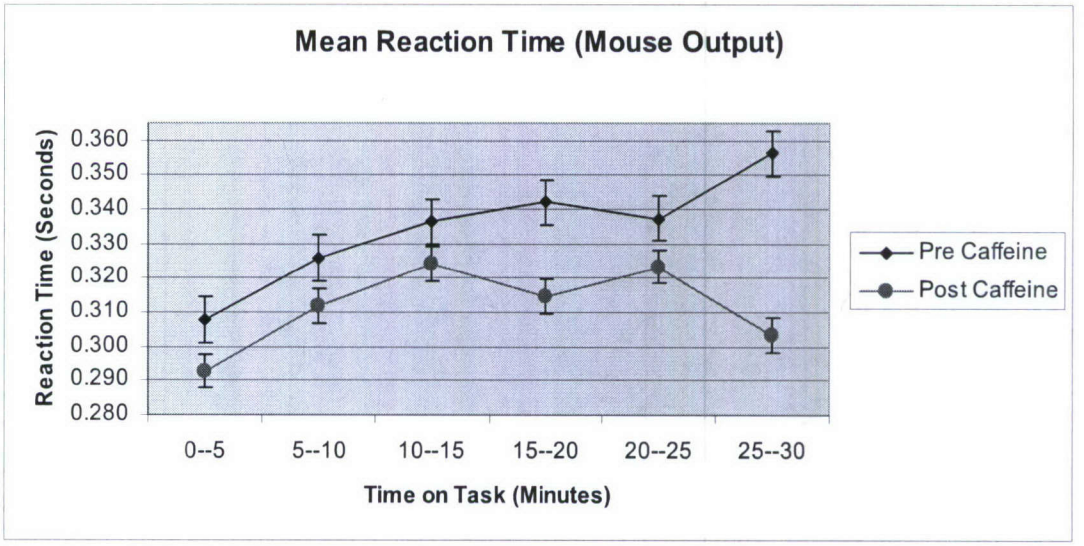


Figure 40: Caffeine Demonstration Mean Reaction Times

Figure 41 shows the median reaction times. As was true in the study described in sections 2.1 and 3.1, the differences between mean and median values can be accounted for by the increase in long latency responses over time.

Figure 42 depicts tonic pupil diameter changes for the pre and post caffeine conditions. Following ingestion of caffeine, the pupils are consistently larger than is true of the pre-caffeine condition. In the pre-caffeine condition, the pupil gets smaller for the first 20 minutes of task performance, it then returns to a level still below that seen at task initiation. In the post-caffeine condition, the pupil gets larger following the initial 5 minutes of task performance and remains at that level for the remainder of the task.

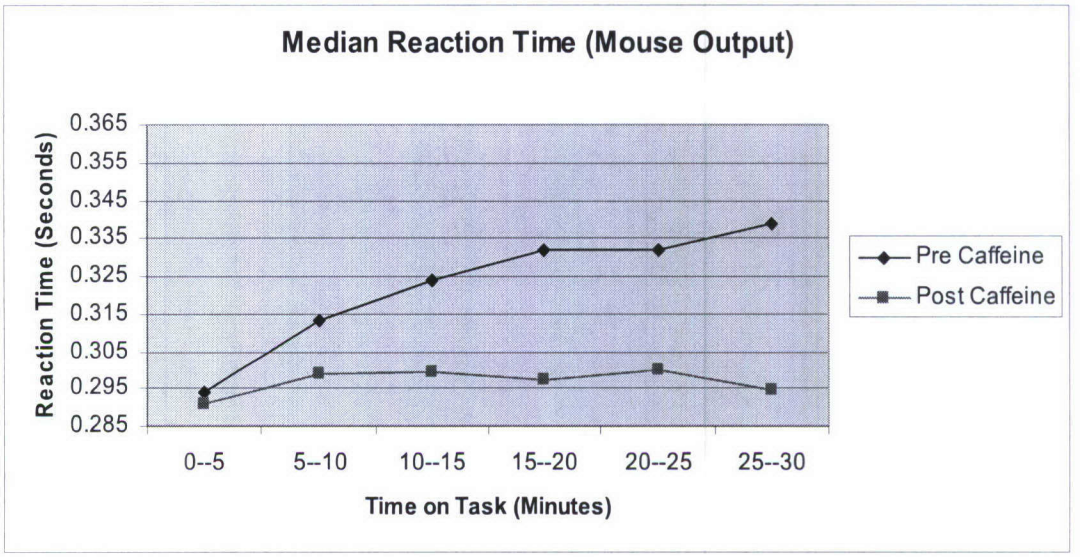


Figure 41: Caffeine Demonstration Median Reaction Times

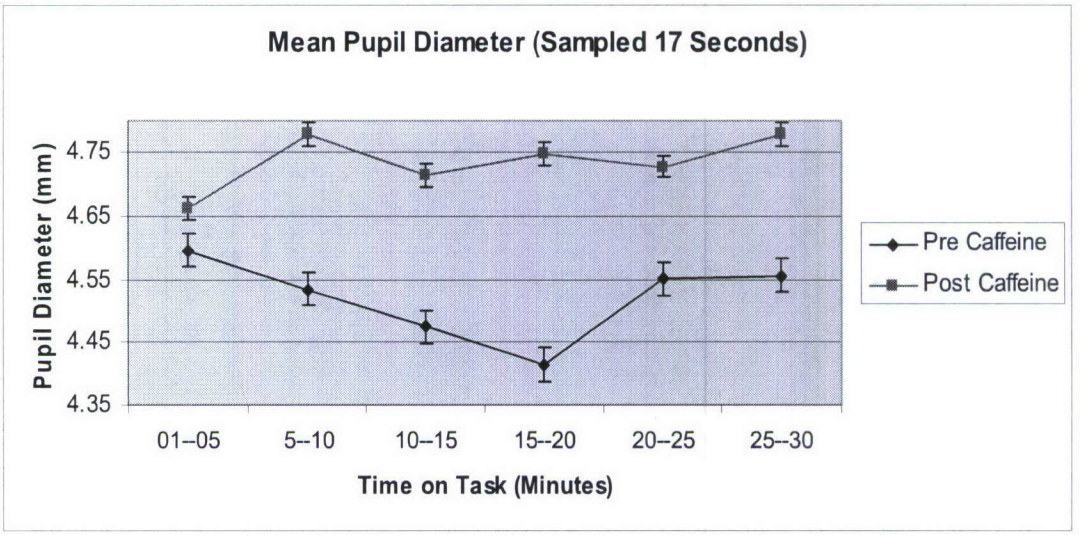


Figure 42: Caffeine Demonstration Mean Pupil Diameter

4 Software Development

As is mentioned in section 1.4 above, we have developed an *Application Framework for Alertness Monitoring Applications* (a.k.a. Alertness Monitor or AM framework) This framework is designed to allow the creation of alertness monitoring applications that are customized for the environment in which the application is to be used. Figure 43 provides an overview of the design of the AM framework.

Creating a software application from a framework is referred to as *instantiating* the framework. In the case of the AM framework, instantiation is done by "plugging" in three customized types of software components: *Event Detectors, Pattern Recognizers,* and a *Lapse Evaluator.*

Event Detectors process the various received channels of data in real-time and detect bio-behaviorally significant events in those streams of data. Such significant events include blinks (or losses of data that result from those blinks), saccades, shifts in pupil diameter, operator responses, pressure changes, changes in heart period, etc. Various event detectors can be plugged-in to the framework to detect events of interest in the current context, one detector for blinks, one detector for saccades, one for shifts in pupil diameter, etc. Each event detector leaves events that it has detected in the *Event Store.*

Pattern Recognizers start their work where the event detectors leave off. That is, they check the Event Store populated by the Event Detectors to look for patterns of events that may indicate, for the current operator/subject, that there has just been or soon will be an alertness lapse. For example, one pattern recognizer might check the event store to determine if there has been a relatively rapid increase in pupil diameter that follows closely on the heels of a subject response. Another pattern recognizer might look for saccadic eye movements with a duration that is longer than predicted based on a regression equation calculated by examining saccades for the subject during an early period of task performance. Another might look for signs of pupillary hyppus, and another, an "eyes closed recognizer", might simply look for signs that the subject/operator has closed her eyes for more than 3 seconds. Each pattern recognizer leaves an indication of any recognized patterns indicative of alertness lapses in the *Lapse Pattern Store.*

The *Lapse Evaluator* checks in the Lapse Pattern Store for sets or combinations of patterns and determines an Indication of Alertness Status to be output by the alertness monitoring application. The indication of alertness status is a simple number ranging from 0 to 100 indicating the application's "judgment" of whether the operator is experiencing a lapse in alertness, with 0 indicating that the application has detected no indication that a lapse is occurring or is about to occur and 100 indicating that the application has detected that a lapse is imminent or currently happening.

For example, one possible algorithm for a Lapse Evaluator is simply to check to see if an "eyes closed recognizer" has left a recognized pattern in the Pattern Store. If such a pattern were found in the store, then this evaluator would output a 100 indicating a lapse is occurring. Another possible algorithm for a Lapse Evaluator is to check to see if any two or more of patterns indicating lapse have been found. Similarly, another possible algorithm would start out with an assumed output value of 0 and add 20 to that value for each lapse pattern found by a pattern recognizer. Once this evaluator determines its output value, this output value can be transmitted

to another application that can determine the appropriate action based upon the current alertness status value.

In Figure 43, solid arrows represent the flow of data through the system. The Data Acquisition Subsystem is an off-the-shelf (OTS) software and hardware system for receiving and digitizing analog channels of data[4]. These channels come from the subject monitoring systems (e.g. an eye tracker or a Laser-Doppler Vibrometer), from our custom hardware (e.g. the pressure sensor enhanced computer mouse), and from the computer system controlling the experiment or with which the subject is working to complete the task at hand. "Send Channels of Data" is a custom component that we have created to capture the data from the Data Acquisition Subsystem and transmit that data in real-time to the Alertness Monitoring Application. The application then stores those channels for processing by the Event Detectors. Each Event Detector reads data from the Channel Store and places detected events in the Event Store. Each Pattern Recognizer reads data from the Event Store and stores pattern recognition results in the Lapse Pattern Store. The Lapse Evaluator then reads data from the Lapse Pattern Store and outputs an Alertness Status.

Dashed arrows in the diagram represent the flow of control as the alertness monitoring application is running. Notice that the control flow is a closed loop very much like a typical real-time control application. The components that are labeled in italics in the figure are those that are to be independently "plugged in" to the AM framework to create an alertness monitoring application.

[4] WinDaq from DATAQ Instruments http://www.dataq.com

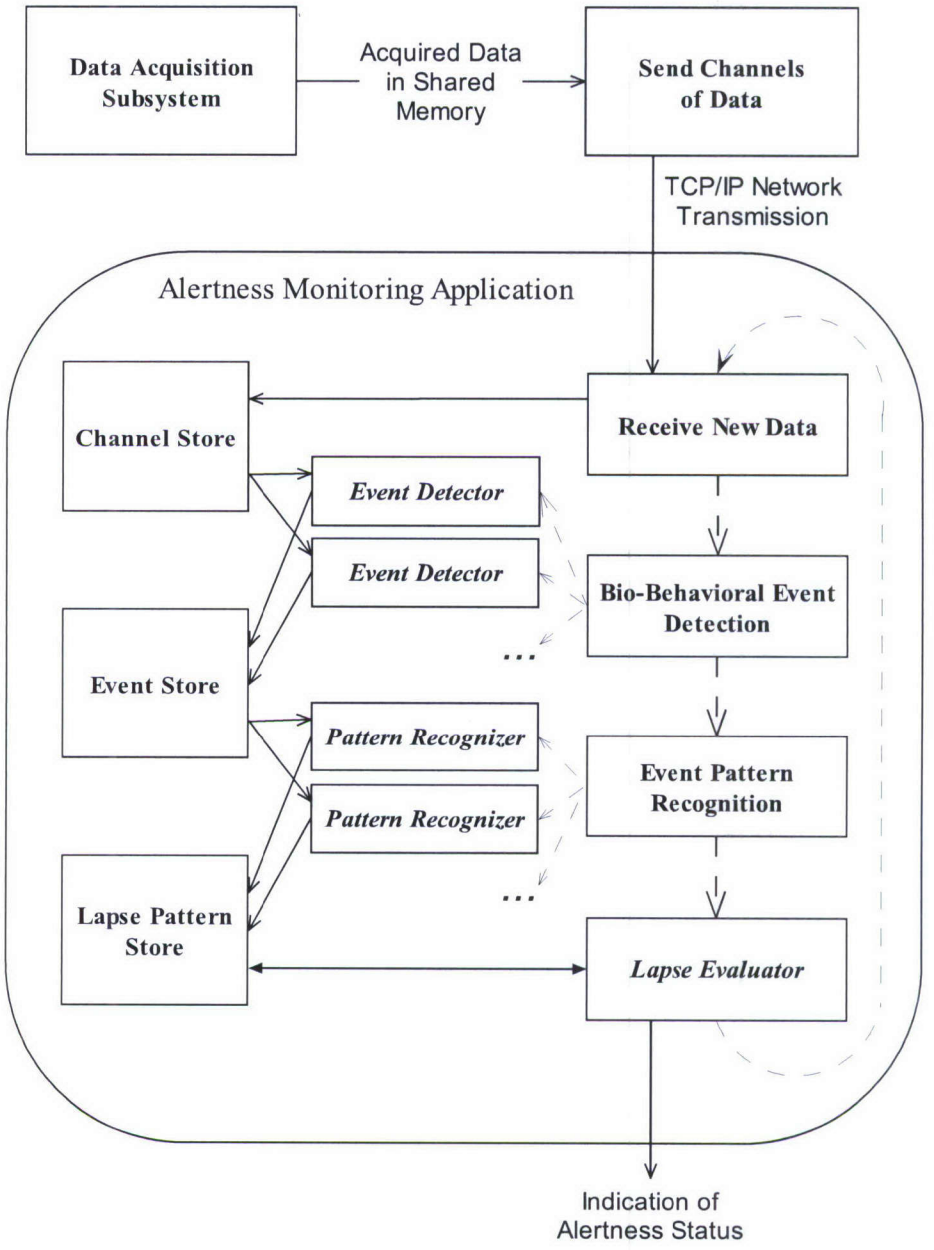


Figure 43: Design Overview of the Alertness Monitor (AM) Framework

5 Conclusions

We have demonstrated that there are both tonic and phasic changes in oculometric variables while subjects perform the EPVT for an extended period. The tonic effects include changes in heart period variability, pupil diameter, and pupil diameter variability as well as some changes in aspects of gaze control such as increased likelihood of long duration saccades. Earlier work identified eye blink variables as also fitting into metric of tonic changes. We also explored the use of response duration as an indicator of tonic as well as phasic changes in alertness. Phasic changes were observed in pupillary responses associated with long latency responses and heart period changes associated with decisions involving the running memory component of the task.

We have developed the software to allow us to monitor most of these events online and algorithms for identifying variables of interest that are reasonably robust. The application of these procedures to identify lapses in alertness is the next necessary step.

As was anticipated, there were marked individual differences in the ability to maintain alertness as well as in the responsiveness of various bio-behavioral measures used to index levels of alertness. *Thus the need for a metric that both monitors a number of bio-behavioral measures and is automatically individualized for the subject being monitored has been confirmed.* We developed such an automatically individualized measure for long duration saccades and believe similar techniques can be applied to other bio-behavioral measures.

We have every reason to believe that practical, unobtrusive, and marketable alertness monitoring systems will result from our work. Therefore, we are actively pursuing funding to continue our development of measures and software to monitor operator alertness status and provide feedback in real-time.

6 References

Bahill, A.T. & Stark, L. (1975) Neurological control of horizontal and vertical components of oblique saccadic eye movements. Mathematical Biosciences, 27, 287-298.

Baker, C.H. (1960). Observing Behavior in a Vigilance Task. Science, 132, 674-675.

Beringer, J. (1992). Timing accuracy of mouse response registration on the IBM microcomputer family. Behavior Research Methods, Instruments, & Computers, 24(3), 486-4990.

Dinges, D.I. & Powel, J.W. (1985). Microcomputer analysis of performance on a portable, simple visual RT task sustained operations. Behavior Research Methods, Instrumentation, and Computers, 17, 652-655.

Donders, F.C. (1868/1969). On the speed of mental processes. Acta Psychologica, 30, 412-431. (Original work published in 1868, republished in Acta Psychologica in 1969.)

McGregor, D.K. & Stern, J.A. (1996). Time on task and blinks: effects on saccade duration. Ergonomics, 39, 649-660.

Plant, R.R., Hammond, N.V. & Whitehouse, T. (2003). How choice of mouse may effect response timing in psychological studies. Behavior Research Methods, Instruments, and Computers, 35(2), 276-284.

Segalowitz, S. & Graves, R.E. (1990) Suitability of the IBM XT, AT, and PS/2 keyboard, mouse, and game port as response devices in reaction time paradigms. Behavior Research Methods, Instruments, & Computers, 22(3), 283-289.

Sogaard, K., Sjorgaard, G., Finsen, L., Olsen, H.B., Christensen, H. (2001) Motor unit activity during stereotyped finger tasks and computer mouse work. Journal of Electromyography and Kinesiology, 11, 197-206.

Ulrich, R., Mattes, S., & Miller, J. (1999). Donder's assumption of pure insertion: an evaluation on the basis of response dynamics. Acta Psychologica, 102, 43-75.

Visser, B., de Looze, M.P., De Graaff, M.P. & van Dieen, J.H. (2004) Effects of precision demands and mental pressure on muscle activation and hand forces in computer mouse tasks. Ergonomics, 47, 202-217.

Wang, L., (1998) Glissadic saccades: a possible measure of vigilance. Ergonomics, 41, 721-732.

Wierwille, W.W., Ellsworth, L.A., Wreggit, S.S., Fairbanks, R.J. Kern, C.L. (1994). Research on vehicle-based driver status/performance monitoring: development, validation, and refinement of algorithms for detection of driver drowsiness. National Highway Traffic Safety Administration Final Report: DOT HS 808 247.

7 Appendix A

The Use of Pupillometric Measures to Detect Moment-to-Moment Lapses in Alertness

Sean D. Kristjansson

John A. Stern

Timothy B. Brown

John W. Rohrbaugh

Washington University

Corresponding author:
Sean D. Kristjansson
Washington University School of Medicine
Department of Psychiatry
4625 Lindell Blvd., Suite 200
St. Louis, MO.
63108, USA.
Ph. 314-286-1358
Fax: 314-454-0432
e-mail: kristjas@psychiatry.wustl.edu

ABSTRACT

Research has shown that small, nonreflexive pupillary changes are robust physiological indicators cognitive activity. In the present paper, we examined whether measures of pupillary changes could be used to detect moment-to-moment (phasic) lapses in alertness during a vigilance task in which eighteen participants responded to a repeated stimulus as quickly as possible. 1.02 second epochs of pupil diameter data following ten long latency responses (indicating low alertness) were compared to 1.02 second epochs following ten normal latency responses (indicating an alert state). A polynomial curve-fitting procedure computed within a multilevel modeling framework indicated that components of the pupil diameter curves following long latency responses significantly differed from those following normal latency responses. Results suggest that components of the pupil diameter curves could be used as neurocognitive markers of operator state in a bio-behavioral alertness monitoring system.

Key Words: Pupillometry, alertness, multilevel modeling

The Use of Pupillometric Measures to Detect Moment-to-Moment Lapses in Alertness

Introduction

Results from field and laboratory studies suggest that human vigilance (close and continuous alertness) degrades rapidly while performing many tasks. Degraded alertness and lapses in alertness can cause accidents in field environments. Recent technological advances in bio-behavioral measurement systems allow for real-time monitoring of physiological and neurocognitive markers of cognitive processing, and for implementing real-time countermeasures to maximize operator performance.

There are several challenges to developing bio-behavioral monitoring systems. Traditionally, physiological and neurocognitive monitoring has required attachment of sensors (e.g., electrodes) to the operator. These inhibit operator movement and limit the number of deployable environments. A preferred approach would be to monitor the operator unobtrusively. Another challenge is known as 'response specificity' (Lacey & Lacey, 1958; Marwitz & Stemmler, 1998); given an identical stimulus, some individuals are more responsive to that stimulus within certain physiological systems whereas other individuals are more response in different physiological systems. Research has shown there are individual differences in with respect to the sensitivity of pupillary changes reflecting aspects of cognitive processing (Beatty & Lucero-Wagoner, 2000). Thus, an ideal system would accommodate individual differences by adapting the system to each individual by combining measures from multiple physiological systems.

Another challenge is that during vigilance tasks there are both tonic changes in alertness as well as moment-to-moment (phasic) changes in alertness. Since both tonic and phasic changes

in alertness can lead to operator errors, a bio-behavioral monitoring system should be sensitive to both types.

Recent work in our laboratory has focused on developing an eye-movement (oculomotor) system capable of real-time monitoring of alertness that could be deployed in many field settings. There are many advantages to developing such an oculomotor-based system. Eye-movements can be monitored remotely in real time. Multiple oculomotor measures can be quantified from camera-based recordings, such as blink frequency, blink duration, eyelid occlusion, saccades (fast gaze shifts), saccade amplitude, saccadic overshoots (fast gaze shifts that miss then return to the visual target), convergence between the eyes as well as pupil dynamics. Research has shown that time on task (i.e., fatigue) manifests as tonic changes in some of these measures (Fukuda, Stern, Brown, & Russo, 2005; Sirevaag, Rohrbaugh, Stern, Vedeniapin, Packingham, & LaJonchere, 1999; Sirevaag & Stern, 2000; Stern, Boyer & Schroeder, 1994).

Pupil diameter change is an informative measure that can be extracted from the eye-camera record. This measure shows promise as a neurocognitive marker for assessing alertness for several reasons. The neurological systems, networks and processes underlying pupil dilation and constriction are well understood (Beatty and Lucero-Wagoner, 2000). Large pupillary changes tend to be driven by reflexes. Two well-known reflexes are primarily controlled by the intensity of light directed at the pupil. The light reflex regulates the amount of light entering the eye, and the accommodation reflex alters foveal focus and decreases binocular disparity. However, smaller, momentary, nonreflexive changes in pupil diameter (up to about 0.5mm) that are unrelated to illumination or accommodation also occur.

Research has shown these nonreflexive changes, known as task-evoked pupillary responses are reliably linked to changes in processing load related to perception, short and long term memory retrieval, preparation and execution of simple movements, language processing, attention and vigilance (for a review see Beatty & Lucero-Wagoner, 2000). Changes in pupil diameter also index tonic changes in cognitive processing due to fatigue (Lowenstein & Lowenfield, 1964; Ludtke, et al., 1998). Further, the pupillary system is a quick-acting system with a high signal-to-noise ratio (i.e., high reliability, Beatty & Lucero-Wagoner, 2000). Consequently, pupillary changes should be ideal for monitoring and detecting moment-to-moment (phasic) changes in alertness during sustained vigilance tasks. Evaluating pupillary changes as indices of alertness levels associated with delayed responses to stimuli (responses that indicate lapses in alertness) compared to 'normal' latency responses (alert responses) during a vigilance task was the primary focus of the present paper.

Specifically, the following research questions were addressed: i) Do measures of pupil diameter changes associated with long latency behavioral responses to stimuli (responses that indicate lapses in alertness) differ from measures of pupil diameter changes associated with normal latency responses to stimuli (responses that indicate the subject was alert)? ii) Are there between-individual differences in the magnitudes of the pupillary changes associated stimuli associated with lapses in alertness compared to pupillary changes associated with alert responses? iii) How early following responding can differences in pupil diameter changes associated with responses that were not alert be differentiated from pupil diameter changes associated with alert responses?

Method

Participants

Participants were twenty undergraduate student volunteers (10 females) and their mean age was 20.0. Data from 2 subjects (1 female) were excluded due to data loss. All participants were healthy with 20/20 vision, corrected or uncorrected. The study protocol was approved in advance by the Institutional Review Board, Washington University. Each subject provided written informed consent and was paid for participating.

Task / Stimuli

Subjects participated in the Enhanced Psychomotor Vigilance Task (EPVT; Stern, Brown & Hodges, 2005). Each stimulus in this task is an incrementing, 4-digit count-up timer presented on the computer screen. The timer starts with all digits set to zero, and subjects are instructed to respond by pressing the left mouse switch as quickly as possible after the timer begins incrementing. The timer stops incrementing at switch closure and the subject's reaction time is displayed. To introduce a running memory component, subjects were instructed to remember whether or not the last digit of their reaction time was odd. When the last digit of three consecutive reaction times was odd (under computer control) the subject is required to press the right mouse switch. Pupillary responses associated with the running memory component of the EPVT were not analyzed in the present study.

Three count-up timers are displayed; one is at the center of the screen and one each 10 degrees to left right of center. Reaction time (RT) feedback was also displayed at these locations. Variation in spatial locus was incorporated to provoke gaze shifts, as is common with most monitoring and control tasks. After feedback was displayed for 400 ms, the timers reset to zero,

and the process repeated. Stimulus onset asynchrony varied from 1.5s to 3.5s from timer reset. Luminance levels of all stimuli and feedback were equated.

Procedure

After agreeing to take part in the study, participants were given written instructions for the EPVT on the computer screen, which were also read aloud, and were provided with a five-minute practice period. After the practice period, subjects participated in the EPVT for 60 minutes.

Apparatus

The experiment was conducted in a dimly lit room (approximately 20 lx). Participants were tested individually, seated in a chair with back of the head comfortably cradled. The headrest contacted the head at two points and allowed for rotational head movements. Stimuli were presented on a computer-controlled display located approximately 57cm in front of the subject.

A video camera located under the CRT display and directed at the subject's right eye recorded eye position, pupil diameter, and pupil occlusion time associated with blinking (LC Technology Eyegaze Development System, Alexandria, VA). The camera system sampled data at 60 Hz. Pupillary changes, eye movements, blinks, stimulus presentation, and manual responses were digitized at a sampling rate of 1000 Hz using Advanced Technologies Data Acquisition System (AT CODAS, DATAQ Instruments, Inc., Akron, OH).

Data editing

A graphic data reduction system, BBDRS, (Bio-behavior Data Reduction System, St. Louis, MO) was used for data editing. Periods of data loss due to blinks were edited using a visual display and a linear interpolation procedure. Each subject's pupillary record was first

edited for blinks using an automated interpolation algorithm. The edits then were scanned visually to identify instances where the automated procedure failed. These blinks were re-interpolated manually.

Selection of pupillary responses associated with lapses in alertness; Long latency responses

For each subject, the 50 longest latency responses (LLRs) were identified from among all the responses made during the EPVT. Over the 60 minute task each subject made in excess of 1000 responses. Response onset was defined as closure of the mouse switch. For each of the 50 LLRs, a 1.02s segment of pupil diameter data was extracted from the digitized data file. 102ms of the pupillary response occurred prior to mouse switch closure (baseline) and the remaining 918ms of the pupillary response followed switch closure. Pupillary responses that comprised 40% or more interpolated data were excluded from the analyses.

Subjects often blinked after responding to stimuli, and the onsets and durations of post-response blinks differed among the subjects. During a blink, the camera could not capture the pupil, which results in data loss. The choice of a 918ms sampling period following responding was based on the observation that most blinks occurred after this time. Consequently, the number of pupillary responses following LLRs that comprised less than 40% interpolated data differed across the subjects. From each subject's remaining pupillary responses, 10 were randomly selected for further analysis.

Selection of pupillary responses associated with alert performance: 'Normal' latency responses

10 pupillary responses following 'normal' latency responses (NLRs) were selected for each subject, so that differences between these and the pupillary responses following LLRs could be evaluated. To better ensure the NLRs represented each subject's 'normal' response times during the EPVT, pupillary responses following the subjects' 50 fastest response times (optimal

performance) were excluded from consideration. An NLR pupillary response was matched to each LLR pupillary response using the following criteria: i) the NLR pupillary response could not comprise 40% or more interpolated data; ii) the stimulus for the NLR had to be presented at the same screen location (left, right, or center) as the stimulus for the LLR; and iii) to control for time on task effects, each NLR pupillary response was sampled as close as possible to the time of LLR onset. If two LLRs were contiguous, then an NLR that occurred preceding the first LLR was selected and an NLR that occurred following the second LLR was selected.

Extraction of measures from the pupillary responses

Twenty pupillary responses (10 following LLRs and 10 following NLRs) were available for analysis for each subject. Several measures were extracted from each of the 20 responses using a multilevel (MLM) polynomial curve-fitting procedure that used maximum likelihood estimation. In the present study, we adapted multilevel modeling procedures described by Raudenbush and Bryk (2002) and Duncan, Duncan and Stryker (2006). For a basic introduction see Kristjansson, Kircher and Webb (2007). This curve-fitting procedure was used because a similar procedure could be implemented into a real- or near-real-time monitoring system.

The raw data used in the curve-fitting procedure were pupil diameter deviations from baseline. Baseline PD was the mean of the 102ms of pupil diameter data prior to response onset. This value was subtracted from each pupil diameter value beginning at response onset through the end of the pupillary response epoch. Since the pupil diameters were sampled at 17ms and the epoch was 918ms, each epoch comprised 54 deviations.

A regression equation was then fit to the 54 pupil diameter values from each epoch for each subject. This was done by regressing, simultaneously, the set of 54 pupil diameter deviations from an epoch onto a set of linear and a set of polynomial coefficients. Thus, the 54

pupil diameter deviations from the epochs were 'smoothed' by fitting a curvilinear regression line (PD curve). Because there were 20 epochs for 18 subjects, there were (20 x 18) = 360 such 'smoothed' PD curves.

The shape of each PD curve was described by a regression equation with three coefficients; a Y-intercept (PD amplitude at a specific position in the epoch), an instantaneous linear slope (PD change rate at a specific position in the epoch), and a quadratic parameter (curvilinear PD change rate across the entire epoch). These coefficients and their variances, covariances and standard errors were computed simultaneously using HLM 6 (Raudenbush, Bryk, Cheong, & Congdon, 2004). The three-level multilevel models (MLMs) used to quantify the measures are described below. The MLMs used to test the research hypotheses are described later.

The first level of the MLM (within-epoch level) comprised the regression equation used to compute each subject's 20 smoothed PD curves:

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(TIME_{ij}) + \pi_{2ij}(TIME^2_{ij}) + e_{tij}$$

w

here,

Y_{tij} was a PD deviation from baseline (mm) at position t (t=*1, 2, 3...54*), in epoch i (i = *1, 2, 3...20*) for subject j (j= *1, 2, 3..18).*

$TIME_{ti}$ was a vector of linear coefficients that indicated the position (t) of each PD deviation in epoch i. Because there were 54 PD deviations beginning at response onset (where t = 1), there were 54 coefficients in this vector. Further, the zero-point in $TIME_{ti}$ was placed between the 30th and 31st samples in each epoch. This zero-point corresponded to approximately 502ms following response onset ($TIME_{ti}$ = -29.5, -28.5, -27.5…21.5, 22.5, 23.5).

$TIME^2_{ti}$ was the square of each of the 54 coefficients in the $TIME_{ti}$ vector (-29.5^2, -28.5^2, -27.5^2…21.5^2, 22.5^2, 23.5^2).

π_{0ij} was the Y-intercept (amplitude) of PD curve from epoch i for subject j at 502ms following response onset. It was computed at 502ms after response onset in epoch i for subject j because the zero-points in the $TIME_{ti}$ and the $TIME^2_{ti}$ vectors were placed between the 30th and 31st sample points in epoch i (see Biesanz, et al., 2004; Wainer, 2000).

π_{1ij} was the instantaneous linear slope of the PD curve at 502ms of epoch i, for subject j. π_{1ij} was the linear slope at precisely 502ms of the PD curve in epoch i because of the zero-point placement in the $TIME^2_{ti}$ vector. When π_{1ij} was positive the pupil was dilating, and when π_{1ij} was negative the pupil was constricting.

π_{2ij} was the curvilinear change rate of the PD curve from epoch i for subject j.

e_{tij} was the residual PD at position t in epoch i for subject j. It was the deviation of the observed PD at position t from the smoothed PD curve. Variance among these residuals was due measurement, sampling, or modeling error.

The differences among each subjects' PD curves were modeled at level 2 of the MLM. Here, 20 Y-intercepts (π_{0ij}), 20 linear slopes (π_{1ij}) and 20 curvilinear change rates (π_{2ij}) were treated as within-subject random dependent variables. Level 2 of the MLM quantified the variability among each subject's 20 PD curves. The level 2 equations were:

$$\pi_{0ij} = \beta_{00j} + r_{0ij}$$
$$\pi_{1ij} = \beta_{10j} + r_{1ij}$$
$$\pi_{2ij} = \beta_{20j} + r_{2ij}$$

where,

β_{00j} was the mean of the Y-intercepts of subject *j*'s 20 PD curves. It was the Y-intercept of subject *j*'s mean PD curve.

β_{10j} was the mean of the linear slopes of subject *j*'s 20 PD curves. It was the linear slope of subject *j*'s mean PD curve.

β_{20j} was the mean of the curvilinear change rates (π_{2ij}) of subject *j*'s 20 PD curves. It was curvilinear change rate of subject *j*'s mean PD curve.

r_{0ij}, r_{1ij}, r_{2ij} were the residual Y-intercepts, linear slopes, and curvilinear change rates for subject *j*. Each residual was the difference between the Y-intercept (β_{00j}), linear slope (β_{10j}), or curvilinear change rate (β_{20j}) from subject *j*'s mean PD curve and the Y-intercepts (π_{0ij}), slopes (π_{1i}), or curvilinear change rates (π_{2ij}) from each of subject *j*'s 20 PD curves. Variances and covariances among these residuals were estimated in the model.

The third level of the MLM was the between-subjects level. The third level quantified the variability among the subjects' PD curves, where coefficients from the 18 subjects' mean PD curves were treated as random between-subjects dependent variables in the level 3 equations:

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{10j} = \gamma_{100} + u_{10j}$$
$$\beta_{20j} = \gamma_{200} + u_{20j}$$

where,

γ_{000} was the mean Y-intercept across all responses and all subjects. It was the sample mean PD Y-intercept at 502ms.

γ_{100} was the mean linear slope across all responses and all subjects.

γ_{200} was the mean curvilinear change rate across all responses and all subjects.

u_{00j}, u_{10j}, u_{20j} were the differences between subject *j*'s mean Y-intercept, mean linear slope, or mean curvilinear change rates and the sample mean Y-intercept, sample mean slope, and the sample mean curvilinear change rate. Variances and covariances among these residuals were estimated in the model.

Once this MLM was computed and the within-subject and between-subjects variances among PD curves quantified, the research hypotheses were tested by adding variables to the level 2 and level 3 equations for Y-intercepts, linear slopes, and curvilinear change rates as shown below.

Results

Dependence of PD curve coefficients on baseline PD levels

Prior to testing the main research hypotheses, the dependence of the subjects' Y-intercepts, linear slopes and curvilinear changes rates on baseline pupil diameters were tested. This was done by entering the 20 baseline pupil diameters for each subject as a predictor variable into the level 2 equations for Y-intercepts, linear slopes, and curvilinear change rates. Note that baseline pupil diameters were 'centered' as discussed in Kristjansson, et al. (2007). This was done to facilitate interpretation of the γ_{000}, γ_{100}, and γ_{200} coefficients in the models used to test hypotheses.

Results indicated that subjects' Y-intercepts and linear slopes were dependent on baseline pupil diameter values, $t(358) = -8.70$, $p < .001$ and $t(358) = -11.85$, $p < .001$ respectively, but curvilinear change rates were not $t(358) = -1.57$, $p = .12$. β_{01j} was -.144 indicating that, on average, 1mm increases in baseline PDs were associated with .144 mm decreases in subjects' Y-intercepts. β_{11j} was -.006 indicating that, on average, 1mm increases in baseline PDs were associated with .006 mm decreases per 17ms in subjects' linear slopes. The results indicated that as baseline PDs increased PD curve amplitudes and linear slopes decreased. We hypothesized that baseline PDs associated with NLRs would be larger compared to baseline PDs associated with LLRs, and this was tested. As expected, subject's mean baseline PDs for NLRs were significantly larger (4.56mm) compared to mean baseline PDs for LLRs (4.24mm), $t(17) = 4.75$, $p < .001$.

Together these results indicated that baseline PDs should be included in the level 2 equations for Y-intercepts and for slopes (but not for curvilinear change rates) while testing hypothesis to control for these dependencies.

Differences between measures of PD curves following LLRs (LLR PD curves) and measures of PD curves following NLRs (NLR PD curves)

Differences among Y-intercepts, linear slopes and curvilinear change rates of the sample mean LLR curve versus the sample mean NLR curve were tested using the MLM defined below. Briefly, a dummy-coded variable (TYPE) that differentiated each subject's 10 LLR PD curves (coded -.5) from the 10 NLR PD curves (coded .5) was entered into the level 2 equations for Y-intercepts, linear slopes and curvilinear change rates. This created a set of coefficients that quantified the differences between LLR and NLR curves, and these coefficients were tested for whether or not they significantly differed from zero. The MLM was:

The level 3 coefficients and results of the statistical tests are shown in Table 1. The coefficients are interpreted and summarized graphically in Figure 1. The sample mean PD curve

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{TIME}_{ij}) + \pi_{2ij}(\text{TIME}^2_{ij}) + e_{tij} \quad \text{Level 1: within-epoch level}$$

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{BASE PD}_{ij}) + \beta_{02j}(\text{TYPE}) + r_{0ij} \quad \text{Level 2: within-subject level}$$

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{BASE PD}_{ij}) + \beta_{12j}(\text{TYPE}) + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + \beta_{22j}(\text{TYPE}) + r_{2ij}$$

$$\beta_{00j} = \gamma_{000} + u_{00j} \quad \text{Level 3: between-subjects level}$$

$$\beta_{01j} = \gamma_{010}$$

$$\beta_{02j} = \gamma_{020}$$

$$\beta_{10j} = \gamma_{100} + u_{10j}$$

$$\beta_{11j} = \gamma_{110}$$

$$\beta_{12j} = \gamma_{120}$$

$$\beta_{20j} = \gamma_{200} + u_{20j}$$

$$\beta_{22j} = \gamma_{220}$$

(grey solid line), the sample mean LLR curve (black solid line), the sample mean NLR curve (black dashed line) all are shown in the Figure.

Table 1. Results of the statistical tests of sample mean Y-intercept, linear slope, curvilinear change rate, and the differences between LLR and NLR Y-intercepts, linear slopes and curvilinear change rates.

Sample Mean PD Curve Parameters	Coefficient	*SE*	df	*t*	*p*
Sample mean Y-intercept (γ_{000})	0.058	.016	17	3.73	< .01
Difference between LLR and NLR Y-intercepts (γ_{020})	0.039	0.0173	357	2.26	< .05
Sample mean linear slope (γ_{100})	0.003	0.0006	17	4.32	< .01
Difference between LLR and NLR linear slopes (γ_{120})	0.0028	0.0005	357	5.91	< .001
Sample mean curvilinear change rate (γ_{200})	-0.000004	0.00001	17	-0.39	ns
Difference between LLR and NLR curvilinear change rates (γ_{220})	0.00006	0.00002	358	3.72	< .001

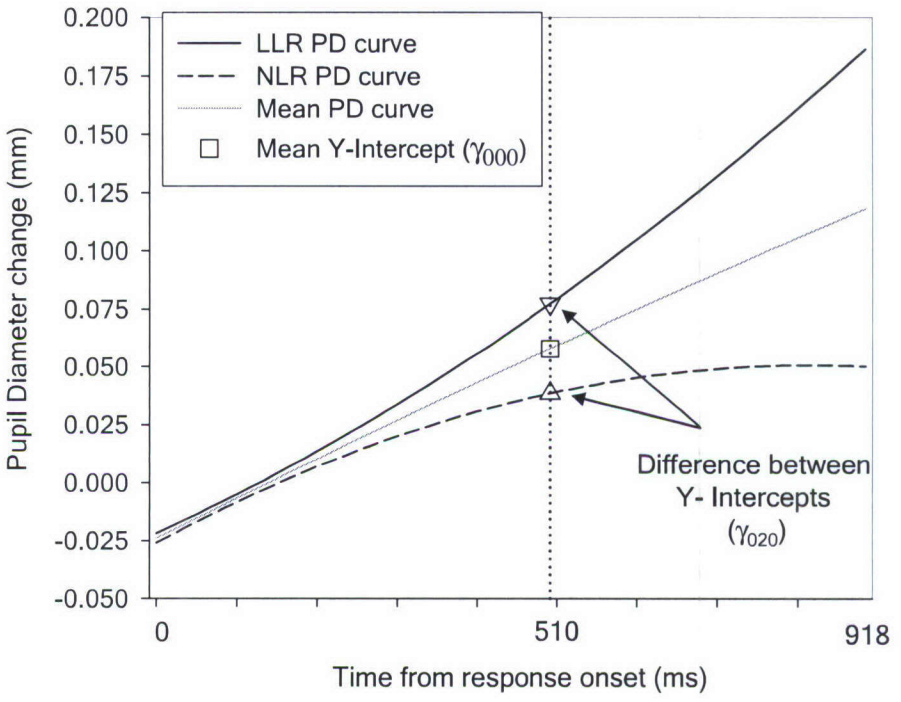


Figure 1. Sample mean LLR and NLR PD curves and sample grand mean PD curve.

The sample mean Y-intercept (γ_{000}) was .058mm, which was the amplitude of the sample mean PD curve after controlling for baseline PD and PD curve type. This is indicated by the black square in Figure 3. At 502ms following response onset, on average, subject's pupils had dilated .058mm above baseline. The sample mean linear slope (γ_{100}) was .003mm. This was linear change rate of the mean PD curve at 502ms across all PD curves after controlling for baseline PD. The sign of γ_{100} was positive indicating that, on average, subject's pupils were dilating at rate of .003mm per 17ms. The sample mean curvilinear change rate (γ_{200}) was -0.000004, and it did not significantly differ from zero. This indicated that when LLR PD curves and NLR PD curves were averaged, there was no evidence of curvilinear PD change.

The difference between the Y-intercept of the sample mean LLR PD curve at 502ms and the sample mean NLR PD curve at 502ms (γ_{020}) was .039mm, and it differed significantly from zero. Results suggested that, as hypothesized, the amplitude of the sample mean LLR PD curve was larger compared to the amplitude of the sample mean NLR PD curve.

The difference between the linear slope of the sample mean NLR PD curve and the linear slope of the sample mean LLR PD curve was .003mm per 17ms, and this parameter differed significantly from zero. This indicated that, as hypothesized, the linear slope of the mean LLR PD curve was larger compared to the linear slope of the mean NLR PD curve. At 502ms, pupillary dilation associated with LLRs was faster .003mm per 17ms faster compared to pupillary dilation associated with NLRs. This difference can be seen in Figure 1; the linear slope of the LLR PD curve (shown within the inverted open triangle) is steeper compared to the linear slope of NLR PD curve (shown within the open triangle).

The difference between the curvilinear change rate of the sample mean LLR PD curve and the curvilinear change rate of the sample mean NLR PD curve was .00006, and this

parameter differed significantly from zero. The curvilinear change rate of the sample mean PD curve was -.000004, which was the midpoint between the curvilinear change rates of the sample mean LLR PD curve and the sample mean NLR PD curve. This meant the curvilinear change rate of the sample mean LLR PD curve was .000026, and the curvilinear change rate of the sample mean NLR PD curve was -.000034. The differences between the curvilinear change rates can be seen in Figure 1.The LLR PD curve in Figure 3 is a 'U-shape' that indicates the linear slope of the LLR PD curve became steeper (moved away from zero) at a rate of .000026mm per 17ms for each 17ms increment across the epoch. In contrast, the NLR PD curve is an inverted 'U-shape' that indicates the linear slope of the sample mean NLR PD curve decreased (moved toward zero) at .000034 mm per 17ms for each 17 increment across the epoch. On average, pupillary dilation associated with LLRs accelerated, whereas pupillary dilation associated with NLRs decelerated.

Between-subjects variability in the size of the differences between LLR PD curves and NLR PD curves

Whether or not there was significant between-subjects variability in the size of the differences between LLR and NLR PD curve parameters was tested; we expected that the differences between LLR PD curves and NLR PD curves would be larger for some subjects compared to other subjects. This hypothesis was tested by adding the residuals (u_{02j}, u_{12j}, u_{22j}) to the level 3 equations for the differences between subjects' mean LLR PD curve Y-intercept, linear slope, and curvilinear change rate and the mean NLR PD curve Y-intercept, linear slope, and curvilinear change rate as shown in the MLM below.

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(\text{TIME}_{ij}) + \pi_{2ij}(\text{TIME}^2_{ij}) + e_{tij}$$ Level 1: within- epoch level

$$\pi_{0ij} = \beta_{00j} + \beta_{01j}(\text{BASE PD}_{ij}) + \beta_{02j}(\text{TYPE}) + r_{0ij}$$ Level 2: within-subject level

$$\pi_{1ij} = \beta_{10j} + \beta_{11j}(\text{BASE PD}_{ij}) + \beta_{12j}(\text{TYPE}) + r_{1ij}$$

$$\pi_{2ij} = \beta_{20j} + \beta_{22j}(\text{TYPE}) + r_{2ij}$$

Level 3: between-subjects level

$$\beta_{00j} = \gamma_{000} + u_{00j}$$
$$\beta_{01j} = \gamma_{010}$$
$$\beta_{02j} = \gamma_{020} + u_{02j}$$
$$\beta_{10j} = \gamma_{100} + u_{10j}$$
$$\beta_{11j} = \gamma_{110}$$
$$\beta_{12j} = \gamma_{120} + u_{12j}$$
$$\beta_{20j} = \gamma_{200} + u_{20j}$$
$$\beta_{22j} = \gamma_{220} + u_{22j}$$

The variances of the residuals VAR(u_{02j}), VAR(u_{12j}), and VAR(u_{22j}) were computed, and tested for statistical significance. Each residual was the difference between the subject-level differences between LLR curves and NLR curves and sample-level difference between the sample mean LLR curve and the sample mean NLR curve. The results are shown in Table 2.

Table 2. Results of significance tests of between-subjects variance components.

Variance Component	Variance estimate	df	χ^2	p
$VAR(u_{02j})$	.0039	17	23.39	ns
$VAR(u_{12j})$	.000004	17	32.52	< .05
$VAR(u_{22j})$	$(.00006)^2$	17	29.25	< .05

Results indicated there was significant variance among the differences between the linear slopes of the subjects' LLR curves versus NLR curves, VAR(u_{12j}), and there was significant variance among the curvilinear change rates of subject's LLR curves versus NLR PD curves, VAR(u_{22j}). However, results indicated that the variance among the differences between the Y-intercepts of subjects' LLR PD curves versus NLR PD curves, VAR(u_{02j}), did not significantly differ from zero; evidence suggested the differences between LLR Y-intercepts and NLR Y-intercepts were constant across the subjects. Consequently, u_{02j} was removed from level 3 equation for β_{02j} for subsequent analyses.

The variability among the linear slopes and curvilinear change rates between subjects' LLR versus NLR curves is shown in Figure 2. The Figure shows that the differences between linear slopes and curvilinear change rates of LLR curves versus NLR curves are larger for some subjects compared to other subjects. For example, the differences between the LLR curve and NLR curve for the subject shown in panel 4 in Figure 2 are large. In contrast, the differences between the LLR curve and the NLR curve for the subject shown in panel 9 of Figure 2 are small.

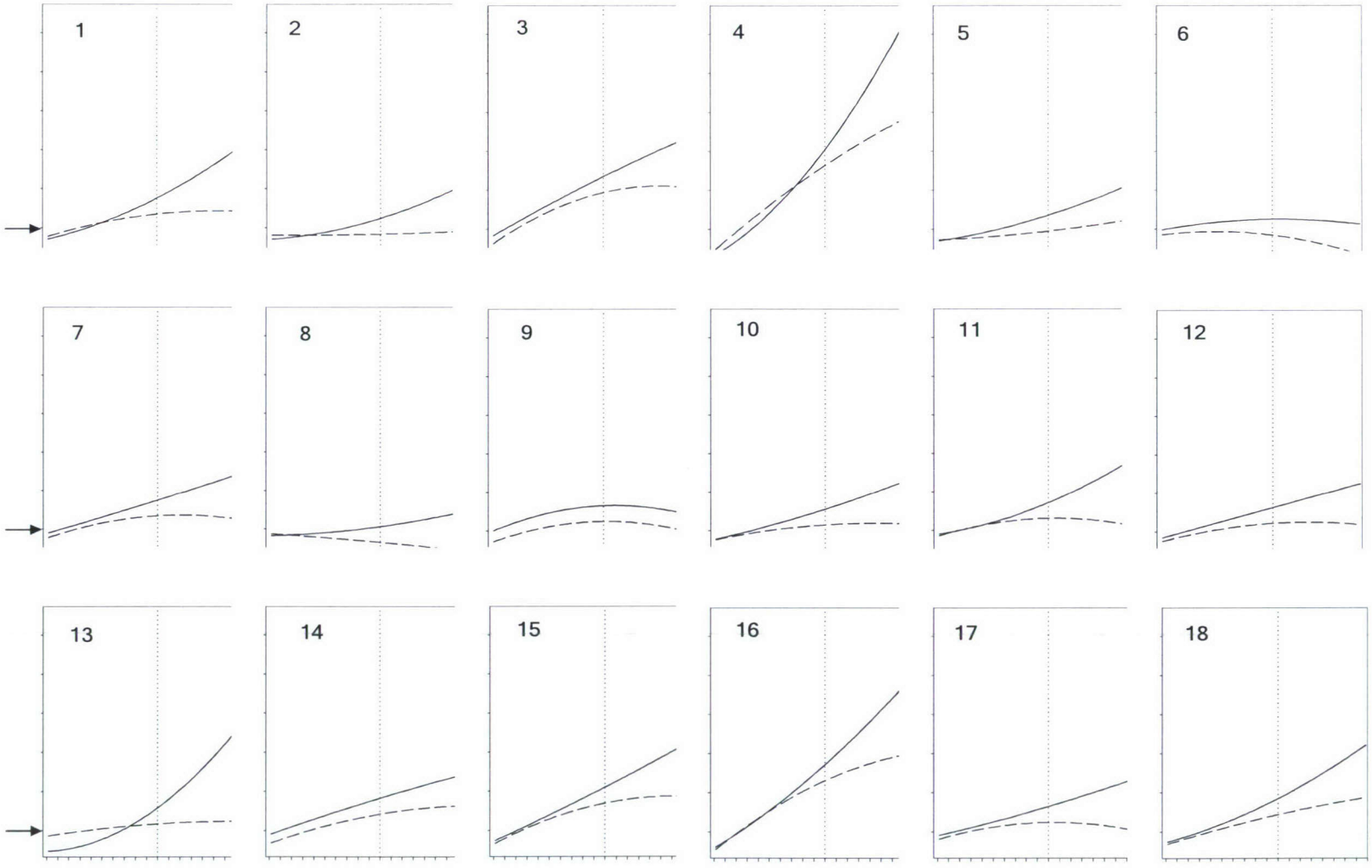


Figure 2. LLR and NLR PD response curves for each of the 18 subjects. The vertical dotted lines indicate the point where Y-intercepts and linear slopes were computed (502ms). The arrow on the Y-axis indicates baseline PD and values on the Y-axis range from -.067mm to .575mm.

We tested if the between-subjects differences in the size of linear slopes and the size of curvilinear change rates of LLR curves versus NLR curves was due to between-subjects differences in EPVT performance; the LLRs for some subjects (e.g. the subject shown in panel 4 of Figure 2) might have been longer for some subjects compared to other subjects (e.g., the subject shown in panel 9 of Figure 2). The mean response latency of each subject's 10 LLRs was computed. The mean of the 18 subject's LLRs was 1.29ms (SD = 0.66). The minimum mean LLR was .45ms and the maximum mean LLR was 2.46ms. Each subject's mean LLR response time then was entered into the level 3 equations for the subjects' differences between linear slopes, β_{12j}, and curvilinear change rates, β_{22j} to test if these differences were due to the different long latency response times.

Results indicated that longer LLRs were associated with larger differences between the linear slopes of subjects' LLR curves versus their NLR curves $t(16) = 5.25$, $p < .001$, and with larger differences between the curvilinear change rates of subjects' LLR versus their NLR curves, $t(16) = 5.05$, $p < .001$. Interestingly, after including the mean of subjects' LLRs in the level 3 equations, the variances of the residuals, $VAR(u_{12j})$ and $VAR(u_{22j})$ did not significantly differ from zero, $\chi^2(16) = 16.34$, $p > .05$ and $\chi^2(16) = 10.58$, $p > .05$, respectively. Essentially all the between-subjects variance in the size of the differences between subjects' LLR linear slopes versus NLR linear slopes, and all the between-subjects variance in the size of the differences between subjects' LLR curvilinear change rates versus NLR curvilinear change rates was due to poorer EPVT performance.

Time points where significant differences between the sample mean LLR PD curve differed from the sample mean NLR PD curve

The time points in the epoch where significant differences between the Y-intercepts and linear slopes of the sample mean LLR PD curve and the sample mean NLR PD curves could be detected was examined by moving the zero-points in the $TIME_{ti}$ and $TIME^2_{ti}$ vectors. Specifically, the zero-points in $TIME_{ti}$ and $TIME^2_{ti}$ were moved in 17ms steps closer to response onset, and the MLM was recomputed at each step. This process was continued until the difference between Y-intercepts and the difference between slopes of the sample mean LLR PD curve compared to the sample mean NLR PD curve no longer differed significantly from zero.

Results indicated that the earliest time point that the Y-intercepts significantly differed was at 450.5ms from response onset. The earliest time point that the linear slopes significantly differed was at 246.5ms from response onset, $t(17) = 1.82$, $p > .05$. These time points are indicated in Figure 3.

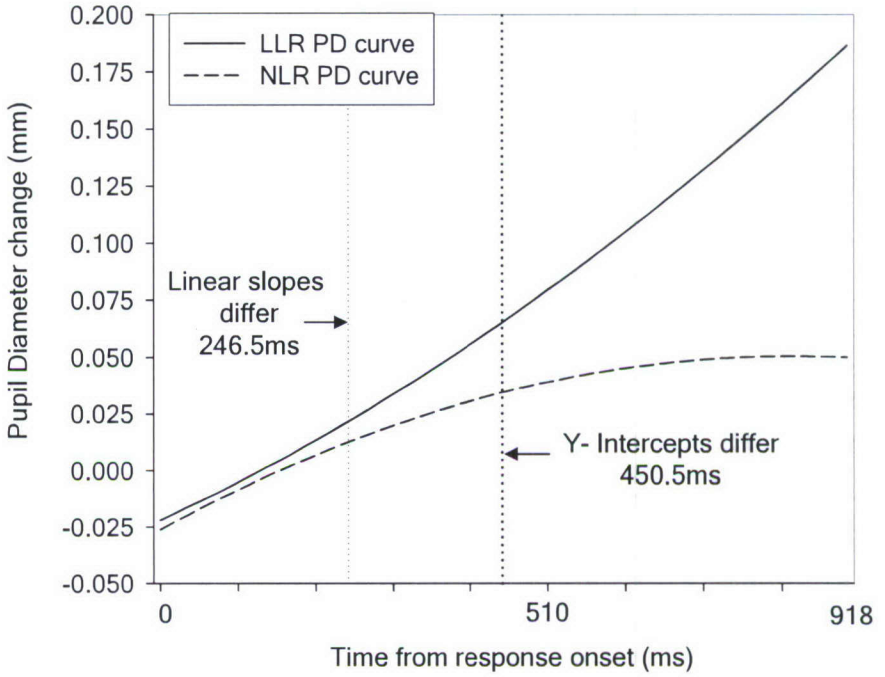


Figure 3. Times from response onset where sample mean LLR and NLR linear slopes and sample mean LLR and NLR Y-intercepts significantly differ.

In summary, baseline pupil diameters preceding LLRs were significantly smaller compared to baseline pupil diameters preceding NLRs. After controlling for baseline pupil diameter, LLR Y-intercepts, linear slopes and curvilinear change rates differed significantly from NLR Y-intercepts, linear slopes and curvilinear change rates. When Y-intercepts and linear slopes were shifted in 17ms steps to earlier points in the epoch and recomputed, the earliest point when LLR Y-intercepts significantly differed from NLR Y-intercepts was at 450.5ms. The earliest point when LLR linear slopes significantly differed from NLR linear slopes was at 246.5ms. Further, there was statistically significant between-subjects variance in the sizes of the differences between LLR Y-intercepts versus NLR Y-intercepts and significant between-subjects variance in the size of the differences between LLR linear slopes versus NLR linear slopes. However, these variances were due to between-subjects differences in EPVT performance; larger differences between LLR curves versus NLR curves were associated longer LLR latencies.

Discussion

Prior research suggests that eye movements and pupillary changes can be used to index tonic, longer-term shifts in alertness and fatigue (Fukuda, et al., 2005; Lowenstein & Lowenfield, 1964; Ludtke, et al., 1998; Sirevaag, et al., 1999; Sirevaag & Stern, 2000; Stern, et al., 1994). The present study examined if pupillary changes can be used to index phasic, moment-to-moment shifts in alertness during a 60 minute EPVT. To examine this question, pupillary measures associated with long latency responses (LLRs) were compared to time-matched pupillary measures associated with normal latency responses (NLRs). The EPVT assumes slow behavioral (long latency) responses to a count-up timer are due to low alertness. In contrast, normal latency responses occur when the subject is alert.

Four pupillary measures reliably discriminated between LLRs and NLRs. One measure was baseline pupil diameter, which was the mean of the 102ms of pupil diameter data collected immediately prior to response onset. Three measures were components of pupillary change following response onset that were computed using a polynomial curve-fitting procedure within a multilevel modeling framework. These measures were pupil diameter change from baseline at 502ms following response onset (Y-intercept; amplitude of the PD curve), the velocity of pupil diameter change at 502ms (instantaneous linear slope of the PD curve), and the curvature of the change in pupil diameter across the 918ms epoch (curvilinear change rate of the PD curve).

On average, the baseline pupil diameter prior to LLRs were 0.32 mm smaller compared baseline pupil diameter prior to NLRs. Pupil diameter change from baseline at 502 ms following LLRs was .039 mm larger compared to pupil diameter change from baseline following NLRs. The velocity of pupil diameter change (dilation) at 502 ms following LLRs was .003 mm per 17 ms faster compared to the velocity of pupil dilation following NLRs. Also, the across-epoch change in dilation following LLRs differed from NLRs. The rate of dilation following LLRs increased continually across the epoch (was a 'U-shape'), whereas the rate of dilation following NLRs increased initially then began to slow (was an inverted 'U-shape').

Together, these results suggest that subjects' level of alertness was lower preceding LLRs compared to NLRs, and the LLR response itself served an alerting function. This is evident because pupillary change following LLRs was a rapid, accelerating dilation that continued across the entire epoch. In contrast, the alerting was smaller following NLRs; the pupil dilated more slowly and began to return toward baseline prior to the epoch termination.

Individual differences

The differences between subjects' LLR PD curves and NLR PD curves were larger for some individuals compared to other individuals. However, these differences were due to EPVT performance. The differences between LLR PD curves and NLR PD curves were smaller for subjects whose LLR times were shorter, and the differences were larger for subjects whose LLR times were longer. After controlling for EPVT performance, between-subjects variance in size of the differences between LLR curves versus NLR curves did not significantly differ from zero.

Toward a bio-behavioral alertness monitoring system

The results suggest that measures of moment-to-moment pupillary changes can be used to index rapid, phasic lapses in alertness. Impressively, changes in alertness can be detected using only 1.02 s of pupillary data, and differences between LLR curves and NLR curves occur within a half second after response onset. Significant differences between LLR pupil diameter changes from baseline and NLR pupil diameter changes from baseline (Y-intercepts of the PD curves) were detected at approximately 450ms following response onset, and significant differences in dilation velocities (differences in linear slopes of the PD curves) were detected at 247ms.

The differences between baseline pupil diameters and the differences between LLR PD curves and NLR PD curves were detected using pupil diameter data from only 10 LLRs and 10 NLRs for 18 subjects. The small number of pupillary responses and the relatively small subject sample size indicate the high signal to noise ratio (high reliability) of pupillary measures as neurocognitive indicators of phasic lapses in alertness. This implies that after a period of system calibration, baseline pupil diameters, Y-intercepts, linear slopes and curvilinear change rates from PD curves fit to pupillary responses in real-time could provide sufficient reliable

information about moment-to-moment alertness so that feedback about operator state would be almost instantaneous.

References

Beatty, J. & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary, and G. G. Bernston (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142-162). New York: Cambridge University Press.

Biesanz, J. C., Deeb-Sossa, N., Papadakis, A. A., Bollen, K. A., & Curran, P. J. (2004). The role of coding time in estimating and interpreting growth curve models. *Psychological Methods, 9,* 30-52.

Duncan, T. E, Duncan, S. E., Stryker, L. A. (2006). *An introduction to latent variable growth curve modeling: Concepts, issues, and applications* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Fukuda, K., Stern, J.A., Brown, T.B. & Russo, M.B. (2005) Cognition, Blinks, Eye Movements, and Pupillary Movements During Performance of a Running Memory Task. *Aviation, Space, and Environmental medicine, 76(7),* C75-C85.

Kristjansson, S. D., Kircher, J. C., & Webb A. K (2007). Multilevel models for repeated measures research designs in psychophysiology: An introduction to growth curve modeling, *Psychophysiology, 44,* 728-736.

Lacey, J. I., & Lacey, B. C. (1958). Verification and extension of the principle of autonomic response-stereotypy. *American Journal of Psychology, 71,* 50-73.

Lowenstein, O. & Loewenfeld, I.E. 1964. The sleep-waking cycle and pupillary activity. *Annals New York Academy of Science, 117,* 142-156.

Lüdtke, H. Wilhelm, B., Adler, M., Schaeffel, F., Wilhelm, H. 1998. Mathematical procedures in data recording and processing of pupillary fatigue waves. *Vision Research, 38,* 2889-2896.

Marwitz, M., & Stemmler, G. (1998). On the status of individual response specificity. *Psychophysiology, 35,* 1-15.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models, applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM6: Hierarchical Linear and Nonlinear Modeling* [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.

Sirevaag, E.J., Rohrbaugh, J.W., Stern, J.A., Vedeniapin, A.B., Packingham, K.D. & LaJonchere, C.M. 1999. Multi-dimensional characterization of operator state: A validation of oculomotor metrics. DOT/FAA/AM-99/28, Office of Aviation Medicine, Washington D.C. 20591.

Sirevaag, E.J. & Stern, J.A. 2000. Ocular measures of fatigue and cognitive factors. In R.W. Backs & W. Boucsein (eds.) *Engineering Psychophysiology.* Lawrence Erlbaum Associates, Mahwah, N.J.

Stern, J.A., Boyer, D. & Schroeder, D. 1994. Blink rate: A possible measure of fatigue. *Human Factors, 36*, 285-297.

Stern, J. A., Brown, T. B., & Hodges, B. J. (2005). A conceptual model for the evaluation of performance lapses. *Bio-behavioral Analysis Systems*, 1-3.

Wainer, H. (2000). Commentary: The centercept: An estimable and meaningful regression parameter. *Psychological Science, 11,* 434-436.

8 Appendix B

Timing of eyeblinks: Random or cognition related?

Authors: Kristjansson, Park, Stern

We have, for a number of years, reported that blinking is not a randomly occurring event but that blinks are most likely to occur in conjunction with the termination of information acquisition and processing. In a study dealing with pupil diameter changes associated with the occurrence of long as compared to normal latency responses we were impressed with the fact that data loss associated principally with blinking appeared to follow a systematic pattern. This report reviews the analysis of that data set. We sampled pupil diameter at stimulus onset, response onset, 340 ms and 1020 ms after responding for trials classified as "long latency" and those classified as "normal latency" RT responses. The following research questions were posed:

Research Questions:

1. Does the number of times that data loss occurs differ as a function of Time (stimulus, response, 340msec after response, and 1020msec after response)?

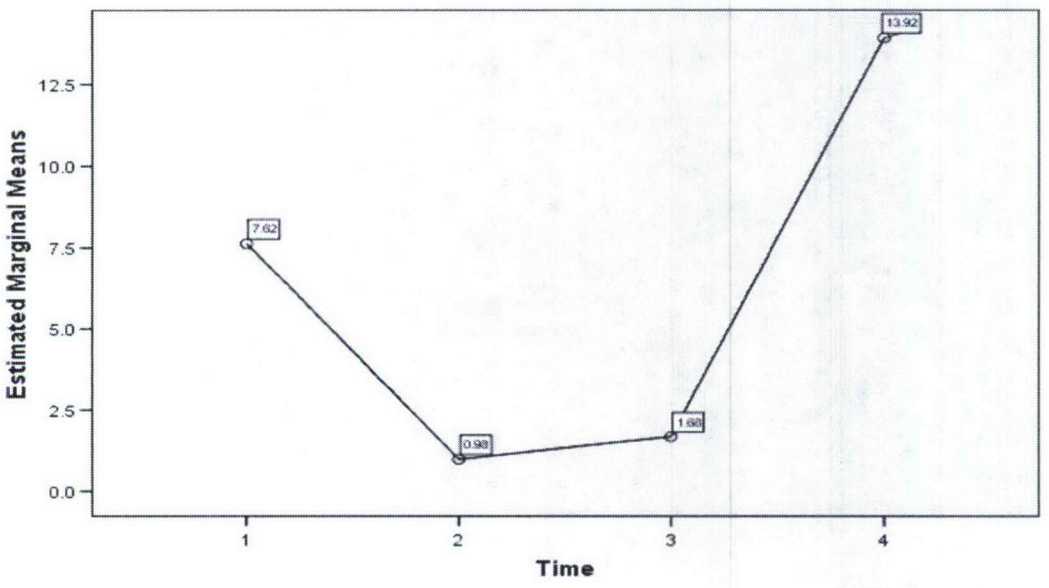
2. Does the number of times that data loss occurs differ as a function of long versus normal response latencies? (Main effect of Latency type)

3. Does the number of times that data loss occurs for long latencies versus normal latencies differ as a function of time period?

Research questions one through three were assessed using a repeated measures analysis of variance (RMANOVA) with 2 within subjects factors: 1) TIME with four levels (stimulus, response, 340ms, and 1020ms); 2) LATENCY TYPE with 2 levels (long versus normal). Significance tests were based on Greenhouse-Geisser adjusted degrees of freedom. Research question 3a was assessed using post hoc paired samples t-tests of means.

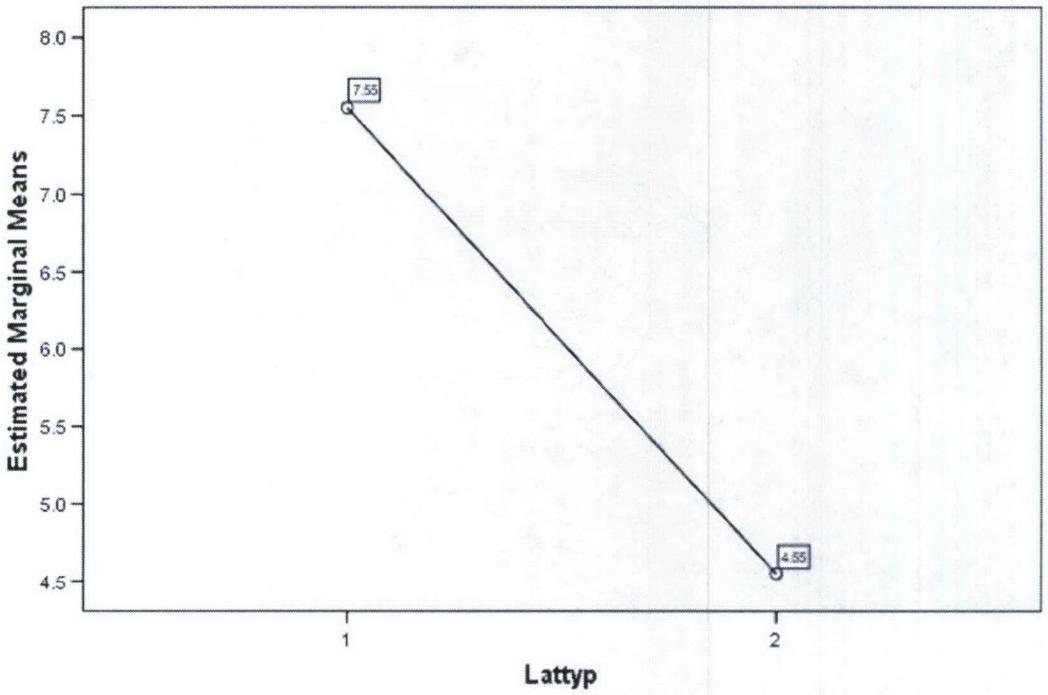
Research Question 1: The test of the main effect of Time assessed for whether or not the number of times that data loss occurred at a particular time period (i.e., at stimulus, at response, at 340ms, and at 1020ms) differed from the grand mean number of data loss occurrences. The grand mean number of data loss occurrences was the mean of the number of data loss occurrences across subjects, TIME periods, and latency types. The main effect of time was significant, $F(1.84, 35.03) = 31.25$, $p < .001$, partial $\eta^2 = .62$. This effect is depicted in Figure 1, below; the number of data loss occurrences is medium, decreases to 340 and then increases drastically at 1020. The largest increase in data loss occurrences occurs at 1020ms.

Figure 1. Main effect of Time.



Research Question 2: The test of the main effect of latency type on frequency of data loss

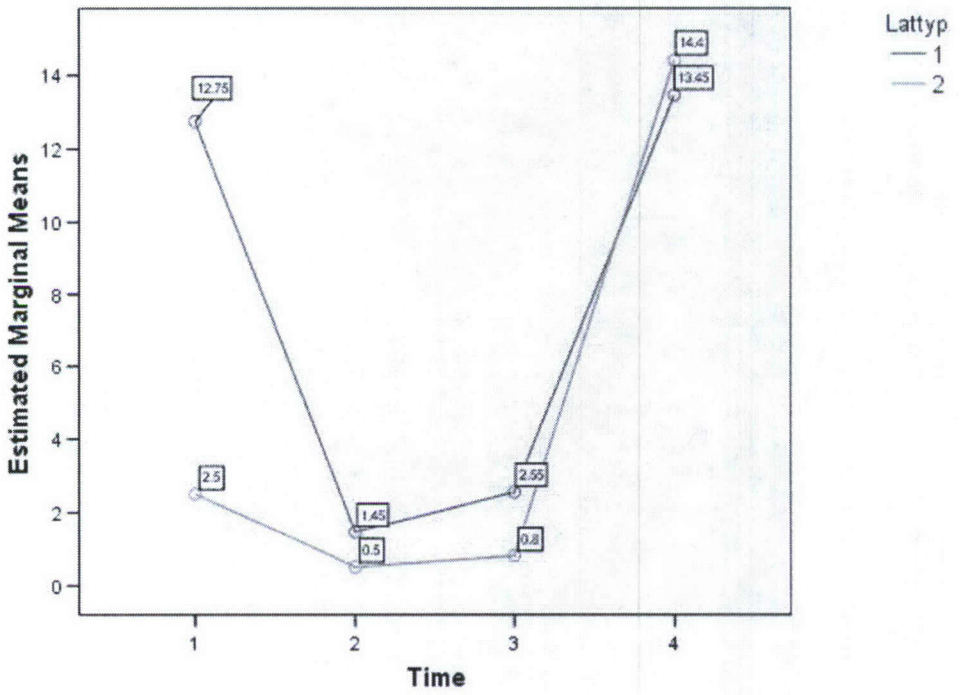
Figure 2. Main Effect of Latency Type.



occurrence assessed for whether or not the mean number of such occurrences for all long latency responses differed from the mean number of data loss occurrences for all normal latency responses. The main effect of Latency Type was significant, $F(1, 19) = 20.13$, $p < .001$, $\eta^2 = .51$ Long latency events generate more periods of data loss than normal latency events. Long latency events generate an average of 7.55, normal latency and average of 4.55 such events

Research Question 3: The test of the interaction between Time and Latency Type assessed for whether or not the mean number of data loss occurrences of long versus normal latencies differed as a function of time. The interaction was significant, $F(1.81, 34.44) = 13.08$, $p < .001$, $\eta^2 = .41$. The means at each time point for both long latency and normal latency responses are shown in Figure 3.

Figure 3. Time X Latency Type Interaction.



Research Question 3(a): Does the mean number of data loss occurrences of long versus normal latency responses at stimulus onset, response, 340ms, and 1020ms differ significantly? Paired samples t-tests were used to test for statistically significant differences between means at each time point. The means of the long latency responses and normal latency responses at each time point are shown in Table 1. Results are shown in Table 2.

Table 1. Number of data loss occurrences means and standard deviations of long and normal latency responses at each time point.

Time point	**Long latency**		**Normal latency**	
	M	***SD***	***M***	***SD***
1. Stimulus	12.75	8.29	2.50	2.95
2. Response	1.45	0.45	0.50	0.18
3. 340ms	2.55	5.07	0.80	1.40
4. 1020ms	13.45	8.62	14.40	2.10

Table 2. Results of paired samples t-tests of pupil diameter means.

Time point	**Mn difference**	**Std Error difference**	**t**	**df**	**p**
1. Stimulus	10.25	1.49	6.86	19	.000
2. Response	.95	0.36	2.65	19	.016
3. 340ms	1.75	1.02	1.72	19	.102
4. 1020ms	-0.95	2.01	-0.47	19	.642

Results indicated that the differences in number of data loss occurrences means at stimulus onset and response onset were significantly different.

Discussion:

The major effect accounting for differences between long latency and normal latency events appears to be data loss at time of stimulus presentation and to a lesser extent at time of response initiation. If the eyes are closed at time of stimulus presentation there is an excellent likelihood that the subject is not going to detect the count-up timer until after the eyes have reopened. Thus the significantly larger number of long latency data loss at time of stimulus presentation can be readily accounted for.

For both normal and long latency responses the increase in blinks at time point 1020 ms after responding is, we believe, associated with termination of the decision making process. The lack of discrimination between LL and NL responses is probably attributable to the fact that data loss duration associated with blinking ranges between 300 – 700 ms. Since the differences in processing time between LL and NL events is generally less than 300 ms the applied metric is not sensitive enough to identify such differences.

9 Appendix C

Effect of task complexity on Reaction Time

Kristjanson, So, Stern

Does the addition of a running memory task affect reaction time to stimulus onset? Donders 1868 in seminal studies demonstrated that as task complexity increased, going from simple to recognition to choice reaction time.

In performing the EPVT the operator has to respond to the presentation of information at one of three possible locations. Following such responding one has to abstract information, namely the speed with which one responded to stimulus presentation and determine whether the number was odd or even. If the number was odd the requirement is to keep track of that fact. If a sequence of three odd numbers occurs the operator is required to make a right mouse response. Since the reaction time component occurs before abstracting information from the display one might expect that reaction time would not be affected by the additional task requirements.

Since gender effects have been described in the literature (Der and Deary 2006;Welford 1980 for example.) we thought it important to evaluate their effect as well as evaluating the order effect with respect to task presentation. Because we were interested in the effect of task complexity on reaction time subjects were required to perform two tasks, one was a recognition reaction time task, the second added the running memory component (or the EPVT with or without the running memory component). Order of presentation was counterbalanced and an equal number (N=6) of each male and female subjects was used.

Appendix C: Effect of Task Complexity on Reaction Time

Method

Participants

The participants were twelve Washington University student volunteers (6 females, 6 males). Participants were required to fill out an informed consent form prior to participating Participants were friends of one of the experimenters (Lapde So).

Materials

A 2 (Order of Task Complexity: Simple First or Complex First) x 2 (Gender of the Participant: Female or Male) between-subjects design was used to explore task complexity effects as a function of gender and order of presentation. Stimulus presentation occurred in a small sound-attenuated chamber. Cameras and an intercom system allow the participant to be seen and heard in the control room. Inside the participant room, a standard computer is positioned on top of a desk where the participant is seated. The participant is provided with a standard computer mouse. Directly outside the participant room is the control room where data is collected. The experiment requires three computers; one for controlling stimulus presentation, the second for abstracting oculomotor data and the third for combining the output from the other two computers. Data is sampled at 1000 Hz.

Procedure

After agreeing to participate in the study, instructions for performing the EPVT were displayed on the computer screen. If the subject did not understand the instructions they were further explained by the experimenter. Subjects then performed the task for approximately 5 minutes or until the experimenter was certain that instructions were understood by the participants. The EPVT presents three count-up millisecond timers that all begin at 0000. One counter is positioned in the center of the screen, one 10 degrees to the left the third ten degrees to

the right of center. Once one of the three timers starts incrementing, the participant is required to press the left mouse button. This stops the timer. In addition, the participant is required to push the right mouse button when three consecutive odd reaction times occur. The frequency of odd terminal integer is, without the participants' awareness, controlled by the experimenter. Approximately 10 such events occur in a 10 minute period. Stimulus onset asynchrony varies from 1.5 seconds to 3.5 seconds. After the participant depresses the left mouse button, the response time remains on the screen for 400 ms. Then, the timer resets to 0000. When the variable stimulus onset asynchrony period ends, one of the three counters begins incrementing and the process repeats itself.

Participants complete two twenty-minute trials with a brief break in between. Break duration was determined by the participant and seldom exceeded 2 minutes.. One trial consists of only the recognition reaction time task (pressing the left mouse button once an incrementing counter is detected;) the other trial is the recognition reaction time task concurrent with the memory component (pressing the left mouse button when they detect a counter incrementing and the right mouse button after three odd values occurred consecutively). The latter is identified as the complex task. Order of stimulus presentation was counterbalanced.

Results:

The data were analyzed using repeated measures analyses of variance (RMANOVA) in which the within-subjects factor was Task Complexity (simple vs. complex) and the between-subjects factors were Gender (Male vs. Female) and Order (Simple task first vs. Complex task first). Statistical tests of the main effects of Task Complexity, Gender and Order, and statistical tests of the Task Complexity X Order and the Task Complexity X Gender interactions were computed. Results are shown in Table 1, and the effects are plotted graphically in Figures 1-5.

Table 1. Results of statistical tests of main effects and interactions.

Within- subjects Effects	**Sum of Squares**	**df**	**Mean Square**	**F**	**p**	**Effect Size (partial η^2)**
Task Complexity	.001785	1	.001785	5.687	**.041**	.387
Task Complexity X Order	.000034	1	.000034	0.097	.763	.059
Task Complexity X Gender	.000600	1	.001000	1.911	.200	.236
Error	.002825	9	.000310	--	--	--
Between-subjects Effects						
Order	.00016	1	.000160	0.055	.821	.006
Gender	.01733	1	0.01733	5.903	**.038**	.396
Error	.02643	9	0.00294	--	--	--

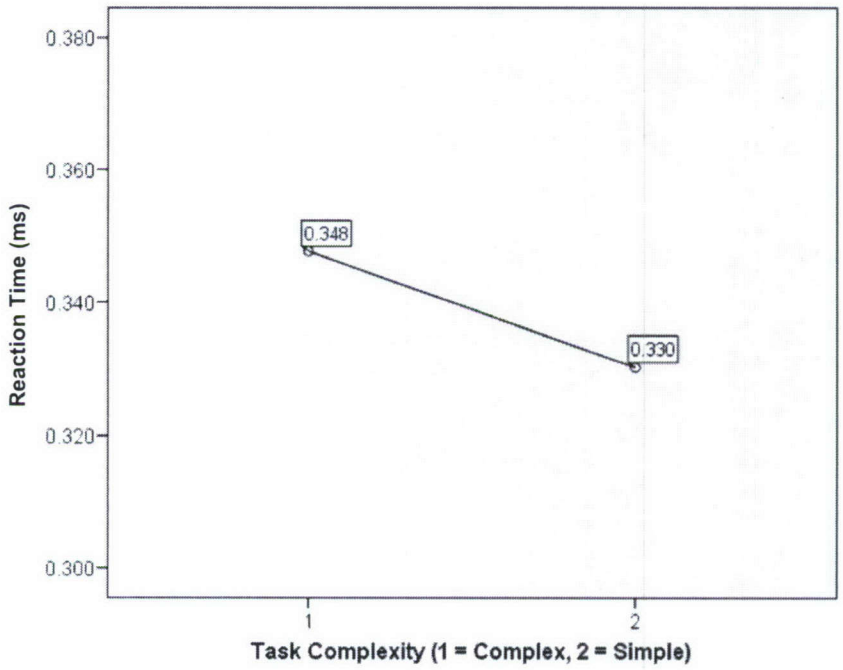


Figure 1. Mean reaction times for complex and simple tasks (main effect of Task Complexity).

The main effect of Task Complexity was statistically significant. Figure 1 shows the mean of the reaction times in the complex task (M = 0.348) was larger compared to the mean of the reaction times in the simple task (M = 0.330). 38.7% of the variance (partial □2) among the subject's reaction times was due to Task Complexity.

The main effect of Gender was also statistically significant. Figure 2 shows the mean reaction time for females was larger (M = .366) compared to the mean of the reaction times for males (M = .312). 39.6% of the variance between subjects' reaction times was due to Gender.

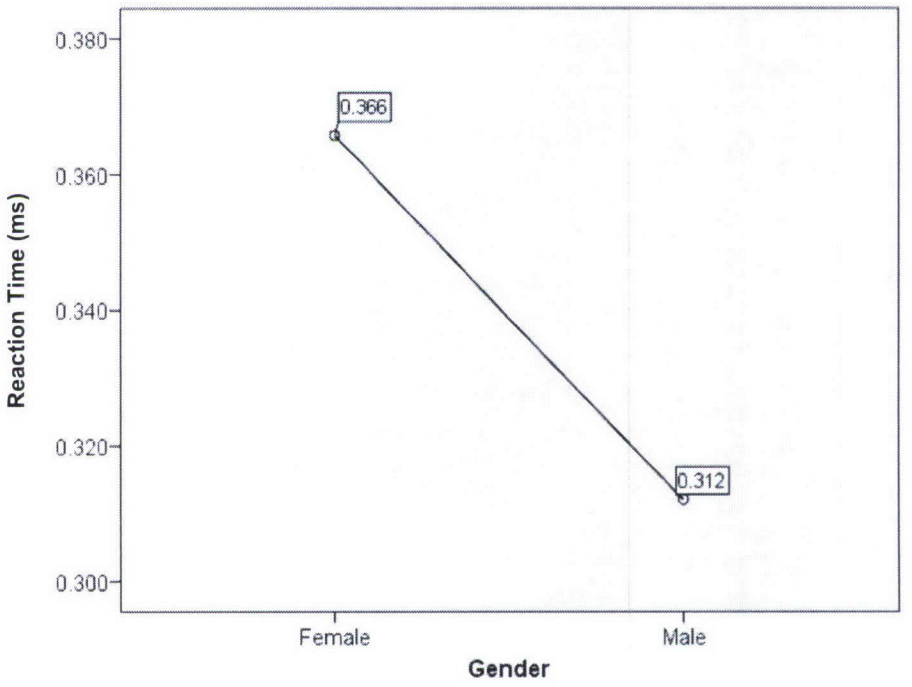


Figure 2. Mean reaction times for females and males combined across complex and simple tasks.

The main effect of Order was not statistically significant. Figure 3 shows the mean reaction time when the complex task was completed first and the mean of the reaction times for when the simple task was completed first. The differences are in the expected direction but not statistically significant.

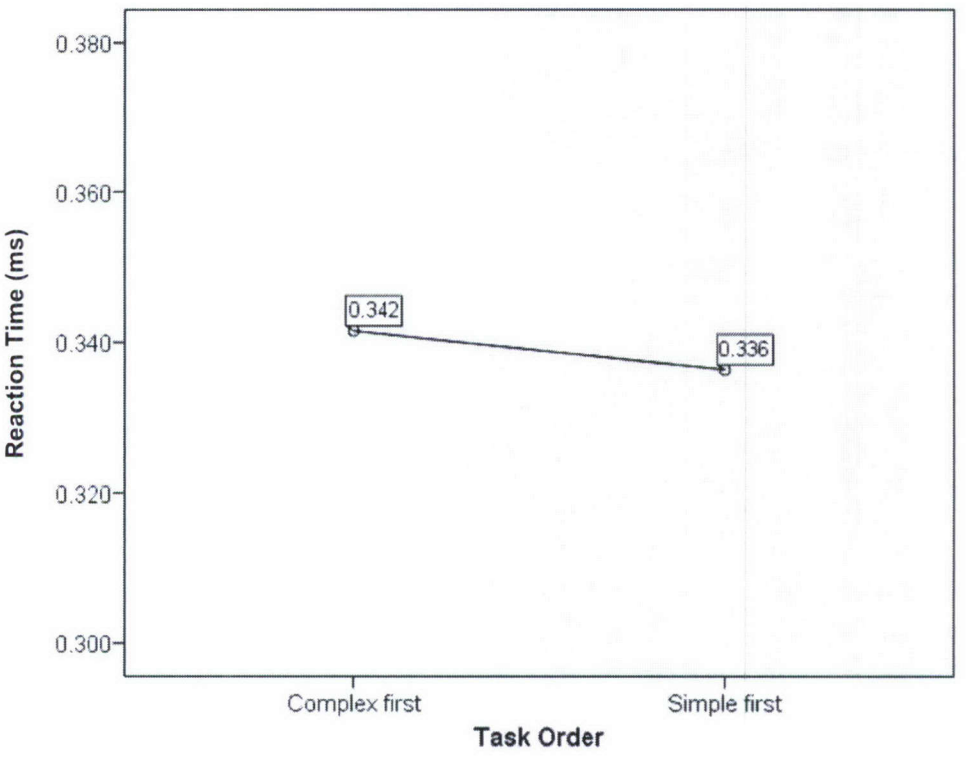


Figure 3. Mean reactions times for each order of presentation.

The Task Complexity X Gender and the Task Complexity X Order interactions are shown in Figures 4 and 5 respectively. Although neither interaction was statistically significant, Figure 5 demonstrates a tendency for the difference between the mean reaction time for complex and simple tasks for females to be larger compared to the difference for males.

These differences may be accounted for by the better performance of females on the complex task.

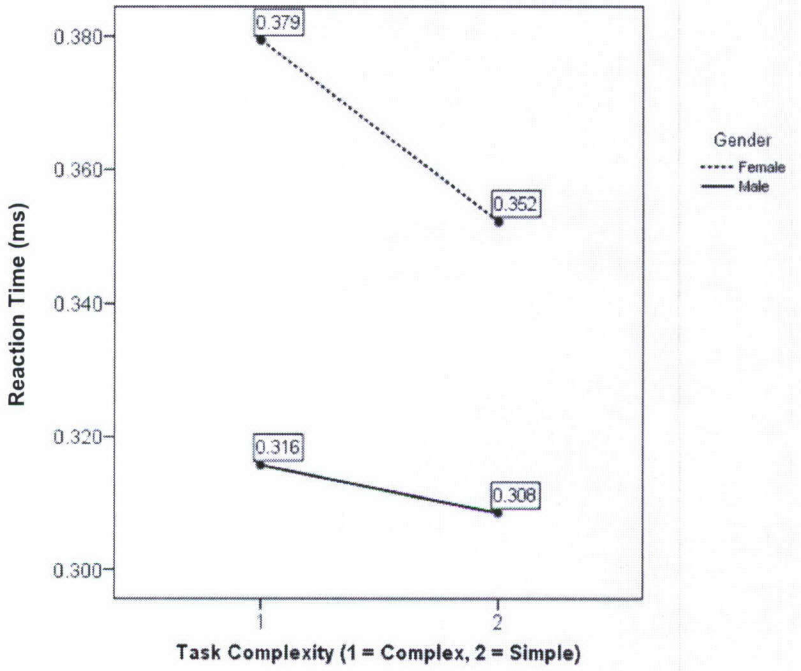


Figure 4. Mean reaction times in complex and simple tasks for females and males.

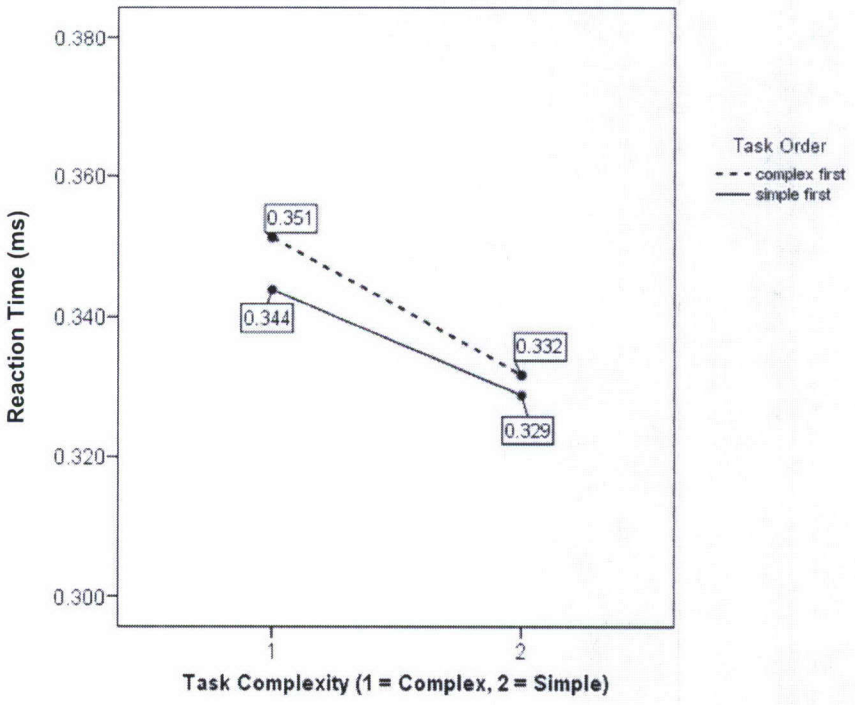


Figure 5. Mean reaction times in complex and simple tasks when the complex task was completed first and the simple task was completed first.

Discussion:

The two significant results dealt with differences as a function of task difficulty and gender. The more difficult task, the one involving a running memory component generates significantly longer RT than the simple RT task. The gender effect found females to be significantly slower than the males. Though there were no significant interaction effects there appears to be suggestive evidence that the difference in reaction time was principally attributable to female subjects demonstrating a disproportionally greater difference between the simple and complex task than was true of male subjects. That task difficulty is related to reaction time has been well demonstrated. For example, in the Der and Deary (2006) study simple reaction time was consistently faster than four choice reaction time. This study, involving data from over 7000 subjects found RT for both tasks to be slower for female than male subjects. Botwinick and Thompson (1966) found reaction times of females also to be significantly slower than was true of males. These authors also measured muscle reaction time, i.e. the initiation of a response, and found that the difference in RT could be principally accounted for by this lag. Our data would suggest a somewhat more complicated picture since RT appears to be differentially affected by task difficulty with the major gender difference attributable to performance on the task involving a running memory component.

Der and Deary (2006) found more consistent differences between male and female groups in RT variability, especially in the coefficient of variation (RT standard deviation/RT mean value), with women being more variable than men. The results could not be interpreted as attributable to speed/accuracy tradeoff. The most critical procedure used to test for this effect sampled only data from subjects who made no errors of the four choice reaction time task and the gender difference remained.

Our results are thus concordant with what one would expect from the literature.

Increasing complexity of a task increases RT. Females respond with longer latencies than males

References

Botwinick and Thompson (1966) Components of reaction time in relation to age and sex. Journal of Genetic Psychology, 108, 175-183.

Der, G. & Deary, I.J. (2006) Age and sex differences in reaction time in adulthood: Results from the United Kingdom Health and Lifestyle Survey. Psychology and Aging, 21, 62-73.

Welford, A.T. (1980) Choice reaction time: Basic concenpts. In A.T. welford (ed) Reaction Times. Academic Press, N.Y. pp 73-128.

10 Appendix D

HEAD MOVEMENTS ASSOCIATED WITH INFORMATION PROCESSING

Chen and Stern

INTRODUCTION TO THE PROBLEM

Head movements:

There is a small but consistent literature suggesting that head movements during information acquisition/processing reflect difficulty of the task for the operator. We first became aware of this possibility some years ago when we observed students at a Junior College enrolled in a remedial reading program. The vast majority of them read using head movements. Observations on older readers reinforce this notion with the additional suggestion that many of these readers not only use head movements in the acquisition of information from text but one frequently sees mouth movements as well, suggesting another variable to consider with respect to evaluating task difficulty. The conclusion about older readers making head movements has to be qualified. Many of them wear bifocal lenses. This may also lead to head movements because near vision is limited to a relatively small area of text.

One can observationally identify when gaze shifts from the end of a line of text to the next line. Gaze shifts associated with right going gaze shifts cannot be so easily identified. If one asks a head moving reader whether they make head movements during reading they will generally not admit to making such movements. After alerting them to the fact that they make such movements they will, when next asked to read, identify that they did make such movements.

Minor head movements have been identified as related to difficulties in the acquisition and processing of textual information. A number of studies with Nechine as the senior author have demonstrated that young readers are most likely to demonstrate such movements and that conceptual difficulty of the text is also an important contributor to such movements. Netchine, Pugh and Guihou (1987) studied 9 and 10 year old readers required to read text in their native language (English and French) as well as in a foreign language they were studying (French and English). Reading in their native language produced relatively little head movements associated with reading while reading in a foreign language produced such movements.

If an adult reader is required to read text aloud one can readily observe such movements in most subjects. The movement most readily observed is return of gaze from the end of a line of text to the next line. Since camera based measures of oculomotor activity can provide information not only about gaze location but also identifies the (theoretical) center of the eyeball, one can use this derived measure to identify head movements. In the current study we manipulated reading difficulty by changing line length as well as changing contrast between the text and background, a purely perceptual manipulation. Our interest was in determining whether changes in difficulty produced by perceptual alterations also produced the changes in gaze control produced by conceptual manipulations.

A review of the research literature predating the Netchine, Solomon and Guihou study found relatively few studies. In a 1924 study on head movement while reading, Fischer reported that such movements were almost always present during reading. Movements to the left were readily apparent while the right going movements are " slow and regular and less easily observed." Fischer also suggested that novice readers make more such head movements than experienced readers. He reported that 11 year olds were likely to show such head movements when reading 6 cm long lines of text while 17-20 year olds only demonstrated such movements with longer line lengths (8-10 cm.) He further reported that reading aloud produced head

movements in all subjects. Fischer reviewed earlier work by Ritzman (1975) and Hering (1879). Hering, holding a cigar in his mouth while reading observed that the cigar moved with his shift of gaze across the text and correctly interpreted this to be associated with head movements during reading. Ritzman had readers clench a pointer between their teeth while reading and reported head movements associated with reading as well as finding marked individual differences in the likelihood of such head movements.

Hardyck and Petrinovich (1970) reported that reading comprehension decreased when oral reading was required, even though reading speed increased. Five years later, Salel and Gabersek (1975) conducted an experiment to determine the influence of head movement on reading performance. They restrained the heads of their subjects and observed decreases in comprehension. They concluded that head movements while reading benefit comprehension.

The current experiment manipulated two variables, namely line length and perceptual difficulty level, with the expectation that both manipulations would make the act of reading more difficult resulting in a higher likelihood of head movements. The fact that line length affects the likelihood of head movements was previously described by Fischer (1924). Netchine and collaborators have demonstrated that cognitive manipulations affect the likelihood of such head movements and we wished, in the current study explore the possibility that perceptual difficulty produced similar effects.

METHODS

Participants:

This study was approved by the Washington University HHSC committee, Six subjects participated, four female and two male; all were between the ages of 16 and 22. Each subject was paid 15 dollars per hour of experimentation, and all subjects contributed useable data to this study. They were instructed that they would be asked two factual questions after reading each article to make sure that they had read and understood the material.

Apparatus:

A text-display program was implemented to control the two variables of interest. This program allows font and background color changes to be made to the displayed text. The colors range from a scale of 0 to 255, with 0 black and 255 white, and gray-scale colors in between. The sequence of text presentation was random.

The Eyegaze Analysis System was used to determine and record gaze activity, including gaze and eyeball center position. The latter was used to monitor head movements. A camera located below the monitor measures gaze activity. The camera samples components of gaze 60 times per second. BBDRS (Bio-Behavioral Data Reduction System), a data analysis program developed by University Software Development, LLC(7), was used to analyze the data.

A two-way ANOVA was used to compute statistical significance.

Procedure:

Material Preparation

1. Nine articles, similar in length and conceptual difficulty were selected.
2. Two questions were generated for each article dealing with factual information contained in the articles. A list of four choices was then produced for each question.

Data Acquisition

1. Eyegaze Analysis System (LC Technologies, Inc.) was used in this experiment to measure and record gaze and head movements.
2. The subject sat in a comfortable chair in front of a computer monitor approximately 25 inches (64cm) from the display.
3. The subject was instructed before the experiment began to try to restrain torso movements as much as possible, and keep the feet planted firmly on the ground.
4. A calibration procedure lasting approximately 30 sec. was conducted.
5. Text was displayed on the monitor, and the subject was instructed to start reading. No time limit was placed on such reading.
6. Each article was displayed on the monitor according to a fixed sequence (See TABLE #2.).
7. The subject used a mouse to control changing the display; all nine articles were more than one page in length. Two questions presented on index cards were shown to the subject after reading each article. The subject was directed to try to answer the questions first from memory (recall). If that was not possible, the subject could then pick an answer from a list of four choices (recognition).
8. The subject then read the next article and answered the two questions for the second article. This process continued until all nine articles were read.

Data analysis

1. BBDRS (Bio-Behavioral Data Reduction System) 5.3.0 (University Software Development, LLC) was used to allow the operator to identify points of interest, specifically initiation time and amplitude of the saccade shifting gaze from the right side of the page to the left as well as time of initiation and amplitude of head movements associated with the line change saccade
2. The data was saved as Microsoft Excel files included: head movement onset and termination latency with respect to the gaze shift onset, the gaze and head movement amplitudes when changing lines, the ratio of the head and gaze amplitude, and the gaze duration per line of text.
3. Descriptive statistics, including the median, were derived from these data.
4. Median was the average value used for the analysis, because of occasional large outliers.
5. The answers to the questions were reviewed. Each subject was scored on a scale of 0 to 18; one point was given for one fully correct answer, 0.5 point was given for not been able to recall, or recalled incorrectly, but was able to recognize the correct answer, and 0 point was given for no answer or answers both recall and recognition parts of the question incorrectly.
6. The scores were arranged according to difficulty level and line length, and tested for significance with ANOVA test.

RESULTS AND DISCUSSION

Head Movement:

Head movement was observed to increase significantly as line length increased ($F(2,10)=9.97, p<.01$). This result is not unexpected since more gaze movements are necessary in order to taking in information presented with greater width. Manipulation of contrast also resulted in significant results ($F(2,10)=4.64, p<.05$) (See FIGURE #1.)

This result suggests that perceptual difficulty whether produced by increasing line length or changing contrast influences head movement amplitude in a similar manner. What other evidence was there that reading the text was more difficult when contrast was decreased?

Time taken to read the line was significantly increased by lowered contrast. These results are depicted in Figure #2. Average gaze duration per line of text increased significantly ($F(2,10)=7.48, p<.05$) as difficulty level increase.

Head/eye movement Ratio:

No evidence was found supporting a relationship between line length and the head to eye movement ratio ($F(2,10)=3.26$,ns). The head and eye movement maintain a fairly constant linear relationship with each other as the number of characters per line increased. However, when perceptual difficulty level increased, the head/gaze ratio increased significantly ($F(2,10)=6.72$, $p<.05$)(See FIGURE #3.).

This analysis of head/gaze ratio provides evidence that as perceptual difficulty level increases, head movement amplitude increases while eye movement amplitude decreases.

Head Movement Onset Latency:

The head movement onset latency with respect to the gaze shift onset (saccade initiation)was calculated and no significant evidence was found relating to either line length ($F(2,10)=2.20$,ns) or difficulty level ($F(2,10)=1.06$,ns). This result disagrees with the observations found in monkeys(2) where, under predictable conditions the head movement is initiated before the saccade. However, experimental conditions were markedly different in these experiments. We inferred that in the case of reading the "operator" can predict when a shift from the end of a line to the beginning of a line is going to occur and that this might lead to head movement initiation preceding the eye movement. We find that for more than 80% of the data, head movements were observed to be initiated concurrent with gaze shift (See TABLE #3.), The other 20% of the data all involved values ranging from –17 to 17 msec. Since the Eyegaze Analysis System takes measures approximately every 17 msec, a change of +/- 17 ms can be considered as due to error of measurement.

One possible reason for the difference in results between those conducted with monkeys is that the conditions differ in these two experiments. In Bizzi's study(2), monkeys were trained to follow the target with their eyes, which demands only perceptual skills. In the present experiment, however, the subjects had to use both conceptual and perceptual skills to perform the task. This difference in tasks may have an effect on the coordination between the head and the eye.

Reading Comprehension

Reading comprehension did not show any effect of line length or difficulty level.

CONCLUSION

This experiment studied the head and eye movements associated with gaze shifts from the end of one line to the beginning of the next line. Head movement was observed to increase significantly as both perceptual difficulty level and line length increased.

The ratio between head movement and saccade was found to increase as difficulty level increases, while line length changes had no significant effect. The head movement is initiated concurrent with saccade initiation, there is no sign of anticipatory head movement as one might expect in situations where the operator can predict when gaze has to shift from the end of a line to the beginning of a new line of text.

TABLE #1 Criteria for Display-Text Program

Characters per Line	Difficulty Level	Lines per Page	Font Size	Align-ment	Back-ground Color	Font Color
50	Easy	38	12	240	100	250
50	Medium	38	12	240	100	120
50	Difficult	38	12	240	100	115
70	Easy	38	12	140	100	250
70	Medium	38	12	140	100	120
70	Difficult	38	12	140	100	115
90	Easy	38	12	30	100	250
90	Medium	38	12	30	100	120
90	Difficult	39	12	30	100	115

Table #1 **Criteria for Display-Text Program** remained the same for each article throughout the experiment. Each article was selected for one of the three options for characters per line based on the length of the article. Difficulty level was chosen randomly for each article.

Note: *this is not the order in which each article was presented during the experiment*

TABLE #2 SEQUENCE OF THE TEXTS

Sequence of the Articles	Characters of Line of Text	Difficulty Level
1	50	Easy
2	70	Medium
3	90	Easy
4	70	Easy
5	70	Difficult
6	50	Medium
7	90	Medium
8	90	Difficult
9	50	Difficult

Table #2 **Sequence of the Text** was the order in which each article was presented to the subjects. This sequence was kept constant for all

TABLE #3 HEAD MOVEMENT ONSET LATENCY RELATIVE TO GAZE ONSET

Characters per Line of Text	50			70			90		
Difficulty Level	Easy	Medium	Difficult	Easy	Medium	Difficult	Easy	Medium	Difficult
Subject #1	0	0	0	0	0	17	0	0	0
Subject #2	0	0	0	0	0	17	0	0	0
Subject #3	-17	0	17	17	0	0	17	17	0
Subject #4	0	0	0	0	0	0	0	0	0
Subject #5	0	0	0	17	0	0	0	17	0
Subject #6	0	0	0	0	0	0	0	0	0

FIGURE #1 HEAD MOVEMENT AMPLITUDE ASSOCIATED WITH DIFFICULTY LEVEL

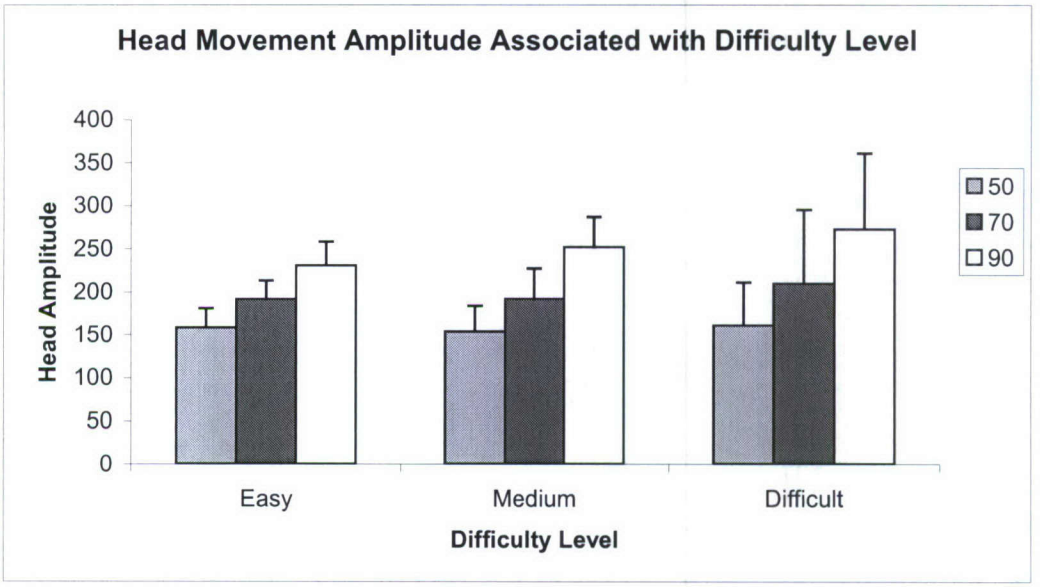


FIGURE #2 GAZE DURATION PER LINE OF TEXT AS DIFFICULTY LEVEL INCREASES

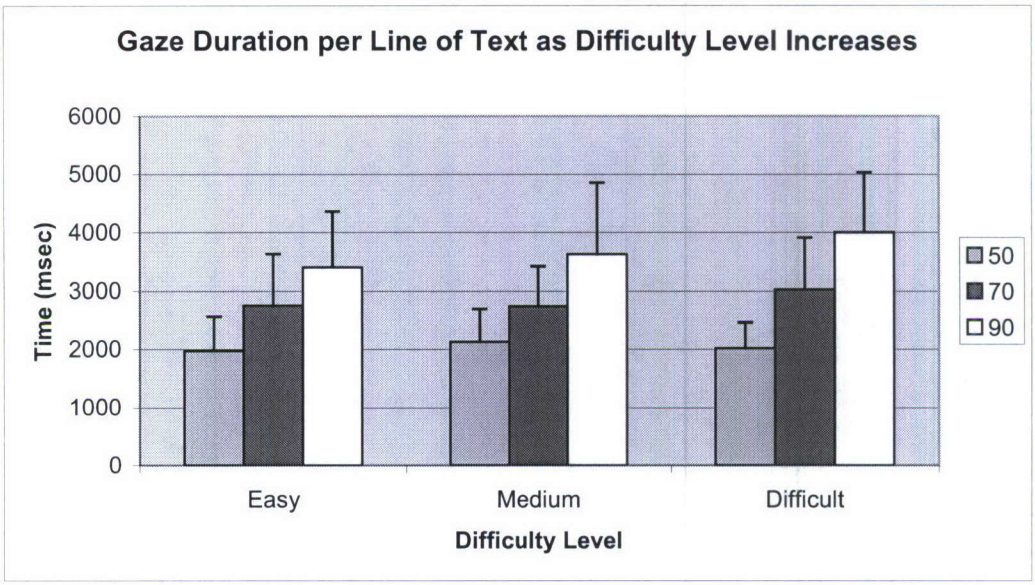


FIGURE #3 HEAD TO GAZE RATIO AS DIFFICULTY LEVEL INCREASES

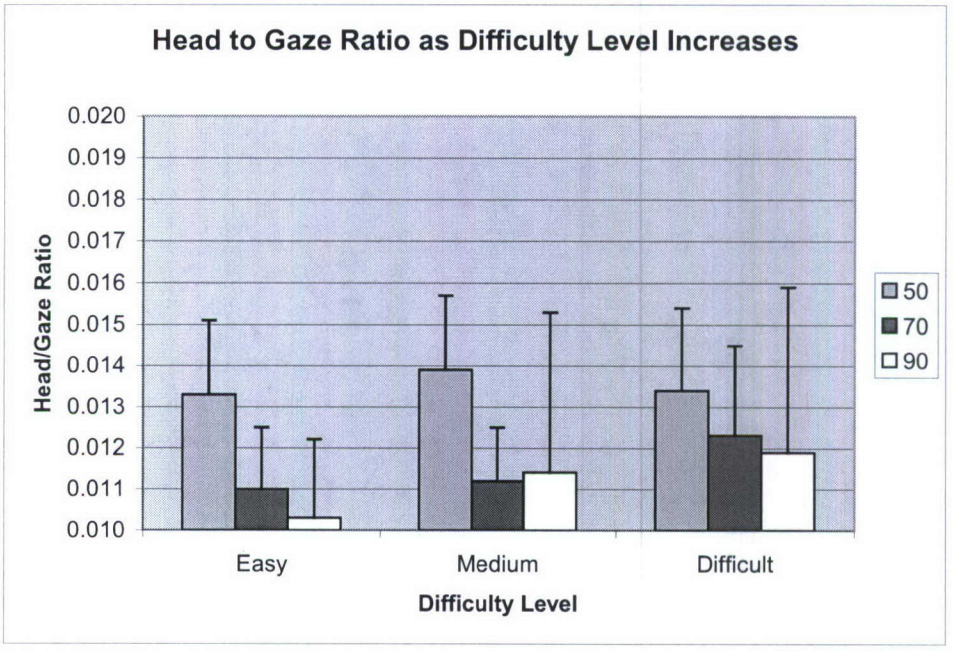


FIGURE #4 BBDRS GRAPHS OF GAZE SHIFTS AND HEAD MOVEMENTS